

VOLUME 39

JANUARY 1961

NUMBER 1

# Canadian Journal of Physics

**Editor: H. E. DUCKWORTH**

**Associate Editors:**

L. G. ELLIOTT, *Atomic Energy of Canada, Ltd., Chalk River*  
J. S. FOSTER, *McGill University*  
G. HERZBERG, *National Research Council of Canada*  
L. LEPRINCE-RINGUET, *Ecole Polytechnique, Paris*  
B. W. SARGENT, *Queen's University*  
G. M. VOLKOFF, *University of British Columbia*  
W. H. WATSON, *University of Toronto*  
G. A. WOONTON, *McGill University*

**Published by THE NATIONAL RESEARCH COUNCIL  
OTTAWA CANADA**

## CANADIAN JOURNAL OF PHYSICS

Under the authority of the Chairman of the Committee of the Privy Council on Scientific and Industrial Research, the National Research Council issues THE CANADIAN JOURNAL OF PHYSICS and five other journals devoted to the publication, in English or French, of the results of original scientific research. Matters of general policy concerning these journals are the responsibility of a joint Editorial Board consisting of: members representing the National Research Council of Canada; the Editors of the Journals; and members representing the Royal Society of Canada and four other scientific societies.

### EDITORIAL BOARD

#### Representatives of the National Research Council

I. McT. Cowan (Chairman), *University of British Columbia*  
Léo Marion, *National Research Council*

H. G. Thode, *McMaster University*  
D. L. Thomson, *McGill University*

#### Editors of the Journals

D. L. Bailey, *University of Toronto*  
T. W. M. Cameron, *Macdonald College*  
F. E. Chase, *Ontario Agricultural College*  
H. E. Duckworth, *McMaster University*

Léo Marion, *National Research Council*  
J. F. Morgan, *Department of National Health and Welfare, Ottawa*  
J. A. F. Stevenson, *University of Western Ontario*

#### Representatives of Societies

D. L. Bailey, *University of Toronto*  
Royal Society of Canada  
T. W. M. Cameron, *Macdonald College*  
Royal Society of Canada  
H. E. Duckworth, *McMaster University*  
Royal Society of Canada  
Canadian Association of Physicists  
P. R. Gendron, *University of Ottawa*  
Chemical Institute of Canada

D. J. Le Roy, *University of Toronto*  
Royal Society of Canada  
J. F. Morgan, *Department of National Health and Welfare, Ottawa*  
Canadian Biochemical Society  
R. G. E. Murray, *University of Western Ontario*  
Canadian Society of Microbiologists  
J. A. F. Stevenson, *University of Western Ontario*  
Canadian Physiological Society

#### Ex officio

Léo Marion (Editor-in-Chief), *National Research Council*  
J. B. Marshall (Administration and Awards), *National Research Council*

*Manuscripts* for publication should be submitted to Dr. H. E. Duckworth, Editor, Canadian Journal of Physics, Hamilton College, McMaster University, Hamilton, Ontario.  
For instructions on preparation of copy, see **NOTES TO CONTRIBUTORS** (back cover).

*Proof, correspondence concerning proof, and orders for reprints* should be sent to the Manager, Editorial Office (Research Journals), Division of Administration and Awards, National Research Council, Ottawa 2, Canada.

*Subscriptions, renewals, requests for single or back numbers, and all remittances* should be sent to Division of Administration and Awards, National Research Council, Ottawa 2, Canada. Remittances should be made payable to the Receiver General of Canada, credit National Research Council.

The journals published, frequency of publication, and subscription prices are:

Canadian Journal of Biochemistry and Physiology	Monthly	\$9.00 a year
Canadian Journal of Botany	Bimonthly	\$6.00 a year
Canadian Journal of Chemistry	Monthly	\$12.00 a year
Canadian Journal of Microbiology	Bimonthly	\$6.00 a year
Canadian Journal of Physics	Monthly	\$9.00 a year
Canadian Journal of Zoology	Bimonthly	\$5.00 a year

The price of regular single numbers of all journals is \$2.00.





VOLUME 39

1961

# Canadian Journal of Physics

**Editor:** H. E. DUCKWORTH

**Associate Editors:**

- L. G. ELLIOTT, *Atomic Energy of Canada, Ltd., Chalk River*
- J. S. FOSTER, *McGill University*
- G. HERZBERG, *National Research Council of Canada*
- L. LEPRINCE-RINGUET, *Ecole Polytechnique, Paris*
- B. W. SARGENT, *Queen's University*
- G. M. VOLKOFF, *University of British Columbia*
- W. H. WATSON, *University of Toronto*
- G. A. WOONTON, *McGill University*

**Published by THE NATIONAL RESEARCH COUNCIL**  
**OTTAWA** **CANADA**



# Canadian Journal of Physics

*Issued by THE NATIONAL RESEARCH COUNCIL OF CANADA*

VOLUME 39

JANUARY 1961

NUMBER 1

## A POLARIZED NEUTRON BEAM PRODUCED BY BRAGG REFLECTION FROM CO-Fe ALLOY<sup>1</sup>

M. A. CLARK AND J. M. ROBSON<sup>2</sup>

### ABSTRACT

A single crystal of Co-Fe alloy is being used to produce a diffracted beam of neutrons with a flux of  $2.6 \times 10^6$  n./sq. cm sec over an area of 1.5 in. by 1.5 in. This beam contains  $92 \pm 5\%$  first-order neutrons ( $\lambda = 1.37 \text{ \AA}$ ,  $E = 0.0436 \text{ ev}$ ),  $3 \pm 5\%$  second- and higher-order neutrons, and  $5\%$  incoherently scattered neutrons. The first-order part of the beam has a polarization of  $0.98 \pm .01$ ; the second-order part of the beam has a polarization of  $0.36 \pm .03$  in the same direction as the first-order part, and the over-all beam has a polarization of  $0.92 \pm .05$ . The neutron spins can be reversed by the magnetic resonance technique with flipping efficiencies of 98% for the first order and 97% for the over-all beam.

### 1. INTRODUCTION

During the last few years the availability of beams of polarized thermal neutrons has made possible the undertaking of many experiments relating to the basic properties of neutrons and to their interaction with matter.\* These beams have been produced either by scattering from or by transmission through magnetic materials; in such beams the resulting polarization arises from the fact that the neutrons not only interact with the nuclei in the material but also, due to their magnetic moments, with the magnetic ion as a whole. This latter interaction depends on the relative orientation of the magnetic moments of the neutron and the ion, and consequently, if the magnetic ions are polarized by an external field, the scattering cross sections differ for those incident neutrons whose spins are parallel or antiparallel to the magnetic field applied to the scatterer. The theory of the magnetic scattering of neutrons has been rather extensively developed (Bloch 1936, 1937; Schwinger 1937; Halpern *et al.* 1939, 1941a, 1941b, 1949a, 1949b; Hamermesh 1942; Steinberg and Wick 1949; Ekstein 1949) and the practical aspect of producing a beam of polarized neutrons has relied on one of three processes: transmission through magnetized iron, total reflection from magnetized mirrors, or Bragg diffraction from magnetized single crystals.

<sup>1</sup>Manuscript received September 28, 1960.

Contribution from the Physics Division, Atomic Energy of Canada Limited.  
Issued as A.E.C.L. No. 1134.

<sup>2</sup>Present Address: Physics Department, University of Ottawa, Ottawa.

\*See, for example, Burgoyne *et al.* 1957; Clark *et al.* 1958; Connor 1959; Dabbs *et al.* 1955b; Nathans *et al.* 1959; Stolovoy 1960; Takei *et al.* 1960; and Trumpy 1957.

The passage of a neutron beam through polycrystalline magnetized iron produces a beam with an excess of neutrons having their spin direction opposite to the direction of the applied magnetic field. The polarization increases with the thickness of the iron since it relies entirely on preferential scattering out of the beam of those neutrons with spin parallel to the field; however, the intensity of the transmitted beam decreases and its relative contamination by fast neutrons and gamma rays increases with increase of iron thickness. The measurement of the polarization of the beam is difficult because it depends rather critically on the velocity distribution of the neutrons striking the iron and on the hardening of the beam by the iron. Rather careful investigations have been made by Hughes *et al.* (1948) and by Stanford *et al.* (1954). The latter group obtained a beam with 32% polarization after transmission through 4 cm of iron. One advantage of the transmission method is that the intensity of the beam is high (Trumy 1957). Clark *et al.* (1958) using a saturated iron block,  $\frac{1}{2}$  in. thick, obtained a beam whose polarization was 24% but whose intensity had only been reduced by a factor of 4 in the transmission; by placing a 3.5-in. quartz filter held at room temperature ahead of the magnetized iron block they lowered the average temperature of the neutrons and thereby increased the polarization to an estimated 27% with a further intensity reduction of a factor of 2. The quartz greatly reduced the contamination of the transmitted beam by fast neutrons and gamma rays (Brockhouse 1959). Another useful aspect of the transmission method of producing polarized neutrons is its simplicity.

The production of polarized neutrons by total reflection from a magnetized mirror was suggested by Halpern (1949a and b) and the use of cobalt was first proposed by Hamermesh (1949). Cobalt is particularly suitable in this application because its magnetic scattering amplitude exceeds its nuclear scattering amplitude and therefore only one spin state is capable of undergoing total reflection. Experimental achievement of a polarized beam by reflection was first achieved by Hughes and Burgy (1951) and was further developed by Burgy *et al.* (1957, 1958). This latter group obtained a beam 5 in. high by  $\frac{1}{4}$  in. wide containing  $7 \times 10^7$  neutrons per second with a polarization which they estimated as  $87 \pm 7\%$ .

Bragg scattering by a ferromagnetic crystal has been extensively used to produce polarized neutrons; most of the early work used the (220) reflection from magnetite (Shull 1951; Stanford *et al.* 1954) but more recently the (111) reflection from face-centered cobalt-iron alloy has been shown to be more suitable (Nathans *et al.* 1959). The theory of Halpern and Johnson (1939) shows that if the incident and diffracted beams lie in a plane at right angles to the direction of magnetization of the crystal then the cross sections for the two neutron spin states (identified as + for the spin parallel to, and - for the spin opposite to, the applied magnetic field) can be written as

$$\sigma+ = 4\pi(b+p)^2$$

and

$$\sigma- = 4\pi(b-p)^2$$

where  $b$  and  $p$  are the nuclear and magnetic scattering amplitudes. Both the (111) and the (200) reflections of Co-Fe have  $p$  almost exactly equal to  $b$  and would therefore be expected to give a very high degree of polarization so long as extinction is not too important. We have used the (200) planes and have achieved a first-order polarization of about 98%.

## 2. THE APPARATUS

The primary beam from the NRU reactor is obtained from a collimator which looks at a rectangular area of about 14 in. wide by 5.3 in. high over which the thermal neutron flux has been estimated (from other measurements not connected with this experiment) to be about  $8 \times 10^{13}$  n/sq. cm sec. The collimator tapers down to a final rectangular aperture of 2.5 in. wide by 1.5 in. high at the outer face of the main shield of the reactor. The measured thermal neutron flux at the center of this aperture is  $1.6 \times 10^{10}$  n/sq. cm sec.

The Co-Fe alloy crystal, mounted between the poles of an electromagnet, is located in this primary beam as shown in Fig. 1 and has been adjusted to

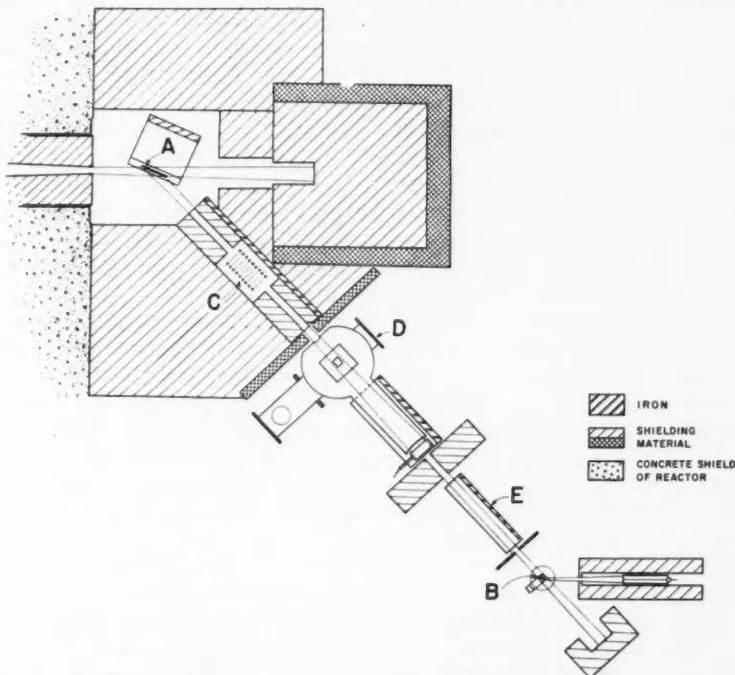


FIG. 1. The experimental apparatus used to produce a polarized neutron beam. A is the polarizing crystal (Co-Fe) mounted between the poles of an electromagnet and located in the primary beam from the NRU reactor. B is the analyzing crystal (Co-Fe) also mounted between the poles of an electromagnet. C is the flipping coil located in a guide magnet, and used to flip the neutron spins through 180°. D and E are the locations of the neutron decay experiment (Clark and Robson 1961) and the capture gamma-ray experiment (Vervier 1960) respectively.

diffract a beam of 1.37 Å neutrons down the second collimator. The rest of the primary beam is stopped in a beam catcher which forms part of the composite shielding system surrounding the Co-Fe alloy crystal and the second collimator. The second collimator is 36 in. long and consists of two parts each with 1.5 in.  $\times$  1.5 in. untapered apertures; it is located between the 36 in.  $\times$  6 in. poles of an electromagnet whose yoke can be seen in cross section in Fig. 1. Between the two parts of this second collimator a radio-frequency coil, 12 in. long by 4 in. in diameter has been arranged with its axis coinciding with the neutron beam; this coil, indicated as C in Fig. 1, can be fed with radio-frequency power to flip the neutron spins by Larmor precession (Stanford *et al.* 1954 and Dabbs *et al.* 1955a). The individual parts of this arrangement will now be discussed in more detail.

Two Co-Fe alloy ingots were obtained from the Virginia Research Corporation\* and various individual crystals were cut with (200) faces.† A slice of another Co-Fe alloy crystal already cut with (200) faces was kindly given to us by Dr. Bozorth of the Bell Telephone Laboratories. Each alloy had sufficient iron to stabilize the crystal in the face-centered cubic form. Preliminary experiments were made with slices from a Virginia ingot and the final assembly was made from the Bell Telephone slice. Comparison experiments indicated that the reflecting powers of both types of crystal are similar. The final assembly consists of two slices of the Bell Telephone crystal each 2 in. by 1½ in. by ½ in. thick with the vertical edges along the [110] direction. The two slices were mounted end to end to form a front (200) face 4 in. by 1½ in. high with the [110] direction vertical. They were adjusted to be parallel by obtaining reflections from a monochromatic neutron beam from both parts and were clamped in this orientation between thin aluminum plates. This assembly was then mounted between the poles of an electromagnet whose pole faces are 12 in. by 1½ in. and are thus much larger than the crystal.

The secondary collimator was lined with cast boron carbide to prevent small angle scattering of neutrons from its inner faces; such scattering would occur with most other forms of inner lining and may lead to depolarization. To maintain the neutron spins vertical during their travel through the collimator a vertical guide field of about 100 gauss is applied by means of the long electromagnet mentioned above. This guide field is, of course, effective on the neutrons as they pass through the radio-frequency coil located between the two halves of the second collimator; the direction of the spin of the neutron can therefore be flipped through 180° by applying radio-frequency power at the Larmor frequency. The radio frequency is obtained from an amplifier driven by a simple LC oscillator. Provided the room temperature does not vary by more than a few degrees the stability of the oscillator and the magnetic field are adequate to maintain the first-order flipping probability within 6% of its optimum value of 98% for periods up to several days without adjustment.

The flux of neutrons at the end of the second collimator has been measured by foil activation to be  $2.6 \times 10^6$  n/sq. cm sec, expressed in terms of a flux

\*Virginia Research Corporation, Richmond, Virginia, U.S.A.

†For cutting this crystal we are indebted to Prof. C. G. Shull of Massachusetts Institute of Technology.

of 2200 m/sec neutrons, and is uniform to within  $\pm 5\%$  across the face of the  $1\frac{1}{2}$  in. by  $1\frac{1}{2}$  in. aperture. The beam then traverses a vacuum tank in which the radioactive decay of polarized neutrons is being investigated (Clark and Robson 1961) and then enters a second guide field. In the region between the two guide fields the spins of the neutrons are held vertical by the leakage field from the magnets; experiments with a dip needle indicated that this leakage field was oriented within  $5^\circ$  of the vertical throughout the volume of interest in the experiment. After passing through the second guide field region the neutron beam is further collimated and is then used for a second experiment in which measurements are made of the circular polarization of the gamma rays resulting from neutron capture (Vervier 1960). A third guide field is applied in this region; the cross section of its yoke can be seen in Fig. 1. The beam is then further collimated and its polarization monitored by diffraction from a second Co-Fe alloy crystal mounted between the poles of an electromagnet and shown as B in Fig. 1. This analyzing crystal was also cut in the (200) plane and oriented with a [110] direction vertical. The doubly diffracted beam is detected by a  $\text{BF}_3$  proportional counter whose axis is parallel to the axis of the primary beam from the reactor.

### 3. MEASUREMENTS OF INTENSITY AND POLARIZATION

#### (a) Intensity

Bacon (1955) has summarized the theoretical considerations by which the intensity of the diffracted beam can be estimated. He expresses the reflectivity of a crystal by means of an angle  $R^\theta$ , called the integrated reflecting power, which can be visualized as a small angular range in the Bragg angle over which the crystal reflects a monoenergetic neutron beam perfectly.  $R^\theta$  is clearly related to  $\eta$ , the mosaic spread of the crystal, and would normally be expected to be of the same order of magnitude; the relation between them, however, is complicated and depends on  $t_0$ , the thickness of the crystal,  $\mu$ , its absorption coefficient,  $\theta$ , the Bragg angle, and  $Q$  where

$$(1) \quad Q = \lambda^3 N_c^2 F_{hkl}^2 / \sin 2\theta.$$

In this expression  $\lambda$  is the wavelength of the neutrons,  $N_c$  is the number of unit cells per cubic centimeter of the crystal, and  $F_{hkl}$  is the structure factor for the  $(hkl)$  reflection. Bacon gives curves relating  $R^\theta/\eta$  to  $Qt_0/\eta \sin \theta$  for various values of  $\mu t_0 / \sin \theta$ . This latter parameter and the factor  $Qt_0/\sin \theta$  can be readily calculated and the curves can then be used to predict  $R^\theta$  for an assumed  $\eta$  or vice versa. For the case of the (200) plane in Co-Fe and  $\theta = 22.7^\circ$  the calculated relationship between  $R^\theta$  and  $\eta$  is shown in Fig. 2.

The reflectivity can also be expressed in terms of a length  $R^\lambda$ , which can be visualized as the wavelength range in an incident white beam over which reflection by a stationary crystal is complete. It is related to  $R^\theta$  by

$$(2) \quad R^\lambda = R^\theta 2d \cos \theta$$

and is an easier quantity to use in estimating the intensity of a diffracted beam.

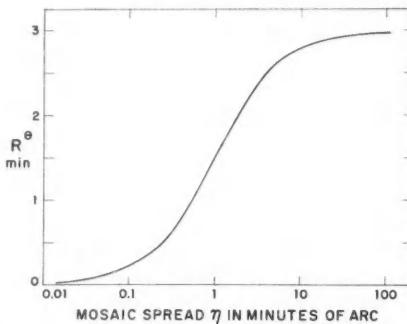


FIG. 2. The calculated variation of  $R^\theta$ , the integrated reflecting power, with  $\eta$ , the mosaic spread, for Bragg diffraction from the (200) planes of a face-centered Co-Fe alloy crystal.

If  $R^\lambda$  is known, it is a relatively straightforward procedure to calculate from the geometry of the experimental arrangement what the diffracted beam intensity should be for a given incident flux of neutrons with a Boltzmann distribution. We have carried out this calculation in reverse and from the measured incident and diffracted intensities have estimated  $R^\lambda$ . From this we have used equation (2) to find that, for our crystals,  $R^\lambda = 0.9$  minute of arc. Figure 2 then shows that the mosaic spread,  $\eta$ , is 0.5 minute of arc.

Direct measurements of the mosaic spread were attempted on a second Co-Fe alloy crystal using the diffracted beam from the first crystal as a source of monoenergetic neutrons. The second crystal was mounted on a turntable located in place of the analyzing crystal in Fig. 1, and standard rocking curves were taken for different collimation angles of the beam incident upon it and of the beam diffracted by it onto the boron trifluoride counter. The results are shown in Table I and in Fig. 3, and show that the measured rocking curve

TABLE I  
Neutron rocking curve widths of Co-Fe alloy crystal  
under different geometric conditions

Geometric collimation angles, ( $2\theta$ )		Observed rocking angle ( $\theta$ ) (minutes)	Curve on Fig. 3
Incident (minutes)	Diffracted (minutes)		
8.6	16.3	13.0	A
8.6	6	5.4	B
4.3	4	3.4	C

width can in each case be accounted for by the collimator angles. Thus these measurements suggest that the true crystal mosaic spread is much less than the smallest collimation angles used; this is therefore consistent with the results of the intensity measurements from which a derived value of 0.5 minute was obtained.

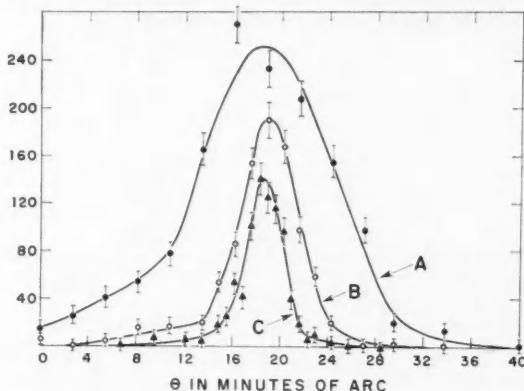


FIG. 3. Rocking curves for a Co-Fe alloy crystal taken with different geometrical collimation angles. The different cases are denoted by A, B, and C and are described in Table I. The ordinates are in arbitrary units and are not directly related for the three curves.

One may well ask how the intensity of the diffracted beam could be increased. Three obvious methods suggest themselves, and are:

- (i) increasing the thickness of the crystal,
- (ii) increasing the mosaic spread of the crystal, and
- (iii) using a multiple crystal.

Increasing the thickness is quite useless due to the high neutron absorption of cobalt. In fact, once  $\eta$  has been determined it is possible to generate a curve relating  $R^\theta$  to the thickness,  $t_0$ , of the crystal. Using our estimated value of  $\eta = 0.5$  minute, the relation is shown in Fig. 4; this shows that there is very little point in increasing the thickness above about 0.05 cm. An experimental check of this was made by cutting slices of a Co-Fe alloy crystal of different thickness and measuring their relative integrated reflecting powers for mono-energetic neutrons. This was done by mounting the samples on a precision goniometer at the location normally occupied by the analyzing crystal; additional finer collimation was introduced so that the beam striking the samples was much smaller than their projected area. Samples of thickness 0.05, 0.1, and 0.3 cm were tried and all gave essentially the same integrated reflectivity in agreement with the predictions of Fig. 4. Thus, apart from mechanical reasons, there is no point in using a crystal more than about 0.05 cm thick.

Increasing the mosaic spread could, in principle, increase  $R^\theta$  and hence the intensity of the diffracted beam. It would appear from Fig. 2 that there is a limit at  $R^\theta$  about 3 minutes; thus increasing the mosaic spread to 10 minutes of arc would result in an increase in the diffracted beam intensity of a factor of about 3 over the value achieved in this experiment.

It is important to note that one must increase the mosaic spread on a microscopic scale and not on a macroscopic scale. Some attempts were made to do this but all were without success. These included bending, bending followed

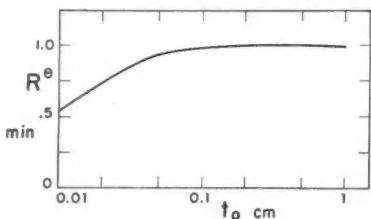


FIG. 4. Calculated values of the integrated reflecting power as a function of crystal thickness for a Co-Fe alloy crystal with a mosaic spread of 0.5 minute of arc.

by flattening, rolling, surface roughening, and surface electropolishing; in each case the reflectivity was measured both before and after annealing but in no case was it increased over the value for the original crystal.

Since the maximum useful thickness of an individual crystal appears to be about 0.05 cm and is therefore less than the mean free absorption path of 0.3 cm, it would appear feasible to make a composite crystal by stacking several thin slices oriented at angles of a few minutes of arc to each other. One might expect to get a small improvement but it will probably be less than a factor of 1.5 because of the oblique path of a neutron in and out of a crystal slice when it is set to reflect neutrons near the maximum of the reactor spectrum.

#### (b) Polarization

The polarization of the beam was measured for its first- and second-order components, with the flipping coil unexcited, by comparing the  $\text{BF}_3$  counter rate with and without a 20-mil mild steel sheet in the beam ahead of the analyzing crystal (Stanford *et al.* 1954). The ratio of these counts, when corrected for background and for absorption by the sheet, is known as the shim ratio,  $S$ , and is related to  $P_1$ , the polarization of the neutrons leaving the polarizing crystal,  $D$ , the depolarization in the guide fields, and  $P_2$  the analyzing efficiency of the analyzer crystal by

$$S = 1 + P_1 D P_2.$$

For this to be valid the sheet must be in a low enough field so that the domains are reasonably randomly oriented and thus cause large enough inhomogeneities in the field within the shim to depolarize the neutrons. Our shim was inserted at a location where the magnetic field was less than 10 gauss.

The measurements on the first-order neutrons ( $\lambda = 1.37 \text{ \AA}$ ) indicated that  $P_1 D P_2$  was  $95.7 \pm 0.7\%$ . The variation of this quantity with the magnetic fields on the polarizing crystal, in the guides, and on the analyzing crystal are shown in Fig. 5. These curves indicate that saturation of the Co-Fe alloy crystal is easily obtained with the apparatus in use and that increasing the guide field above 50 gauss has very little effect on the resultant value of  $P_1 D P_2$ . We have interpreted this last observation as an indication that  $D$  was unity within the errors of the observation since otherwise we would expect  $P_1 D P_2$  to increase as the guide field increases. Further confirmation of this

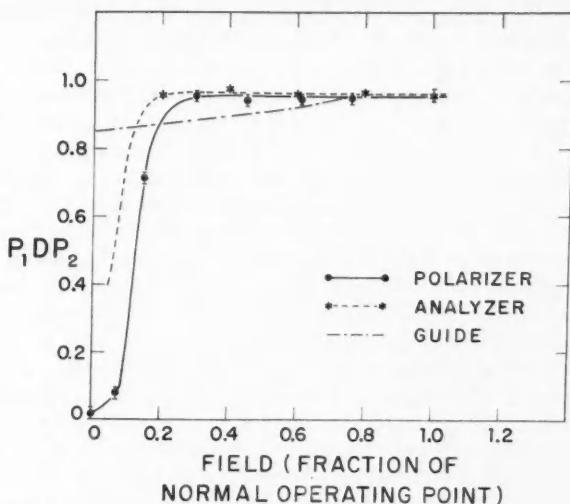


FIG. 5. The product  $P_1D P_2$  plotted as a function of the field on the polarizing crystal, the field on the analyzing crystal, and the guide field. The abscissae are expressed as fractions of the normal operating points, which are 4100 gauss in the open gap of the magnet for the polarizing crystal, 2800 gauss in the gap of the analyzing magnet, and 62 gauss at the beam position in the guide magnet. For each individual curve the other two fields were set at their normal operating points.

was obtained from the fact that a very similar value of  $P_1D P_2$  was obtained in a separate measurement with only half the path length and thus half the length of guide field between polarizer and analyzer. We thus put  $D = 1.00 \pm .01$ . This gives

$$P_1 P_2 = 0.96 \pm .01$$

and if, based on the saturation of  $P_1D P_2$  with magnetic fields on the crystals and on the similar nature of the polarizing and analyzing crystals, we further assume that  $P_1 = P_2$ , then the polarization of the first-order component of the beam is

$$P(1) = 0.98 \pm .01.$$

The effect of second-order neutrons on this result is estimated to be less than the experimental error.

The polarization of those neutrons diffracted in second order by the polarizing crystal was measured by adjusting the analyzing crystal and  $\text{BF}_3$  counter to diffract neutrons of  $0.685 \text{ \AA}$ . The analyzing crystal was then diffracting in its first order those neutrons which had been diffracted in second order by the polarizing crystal. Shim ratio measurements under these conditions gave  $P_1D P_2 = 0.35 \pm .03$ ; assuming  $D = 1$  and  $P_2 = 0.98$  this gives for the polarization of the second-order neutrons:

$$P(2) = 0.36 \pm .03.$$

The polarization of neutrons diffusely scattered from magnetized Co-Fe alloy was estimated for 1.37-Å neutrons by using the shim technique with the polarizing crystal set to diffract first-order neutrons but with the analyzing crystal rocked well off its Bragg curve. The analyzer was therefore deflecting the 1.37-Å neutrons onto the  $\text{BF}_3$  counter by diffuse scattering. Measurements were made for various currents through the magnet on the analyzing crystal and indicated that  $S' - 1 = 0.21 \pm 0.06$  at the normal operating current. If we assume that the shim ratio is related as before this would give  $P = 0.22 \pm 0.06$  for the polarization of diffusely scattered 1.37-Å neutrons. The average value of the diffuse polarization over the pile spectrum was then calculated from this estimated value at 1.37 Å by using the variation of the magnetic scattering amplitude  $p(\lambda)$  with the form factor  $F(\lambda)$  at a fixed incidence angle  $\theta$  to the crystal lattice. The magnetic form factors measured by Nathans and Paoletti (1959) were used in the calculation. This gave  $P(3) = 0.20 \pm 0.06$ .

To obtain the over-all polarization of the beam it was necessary to know the number of second-order neutrons and the number of neutrons in the beam which were not Bragg reflected. The number of diffusely scattered neutrons was measured by rocking the polarizing crystal away from its Bragg peak and comparing the beam intensity measured with a fission counter mounted at the end of the second guide field. This indicated that the diffusely scattered neutrons amounted to  $4.48 \pm 0.05\%$  of the total beam.

The number of second-order neutrons was measured by the absorption technique using gold as the absorber. Fission counter rates were taken both on and off the Bragg peak with gold and with boral absorbers and were combined to give the fraction of second-order beam as

$$a_2 = (3 \pm 5)\%.$$

The main source of error in this measurement is the uncertainty in the total cross section of gold.

The composition of the beam has been summarized in Table II, which shows that the over-all polarization is  $.92 \pm .05$  in the unflipped condition.

TABLE II  
Beam composition and polarization in the flipped and unflipped conditions

Order	Abundance	Polarization		Flipping probability
		r-f. off	r-f. on	
First	$.92 \pm .05$	$.98 \pm .01$	$-.94 \pm .02$	$.98 \pm .01$
Second	$.03 \pm .05$	$.36 \pm .03$	$-.01 \pm .04$	$.49 \pm .01$
Diffuse	$.05 \pm .01$	$.20 \pm .06$	0	$\sim 0.5$
Whole beam	1	$.92 \pm .05$	$-.86 \pm .05$	$.97 \pm .03$

#### 4. MEASUREMENTS OF THE FLIPPING PROBABILITIES

We define the flipping probability  $f_1$  for the first-order neutrons as the probability that an individual neutron will have its spin direction inverted after passage through the r-f. flipping coil. With this definition it is easy to

show that the ratio of the polarization of the first-order neutrons with the r-f. on to their polarization with the r-f. off is

$$P_1(\text{on})/P_1(\text{off}) = 1 - 2f_1.$$

We apply a similar definition for  $f_2$ , the flipping probability for the second-order neutrons.

The flipping probabilities were measured by obtaining the ratio of the counting rate of the  $\text{BF}_3$  counter with the r-f. on to that with the r-f. off. This ratio, which we will call  $r_1$ , if the analyzing crystal is set to reflect first-order neutrons, is related to the flipping probability by

$$f_1 = (1 - r_1)(1 + A_1)/2A_1$$

where  $A_1 = P_1 D P_2$  for the first-order neutrons. Using this method the flipping probability was measured for the first-order neutrons as a function of the current through the flipping coil and is shown in Fig. 6. At each measurement

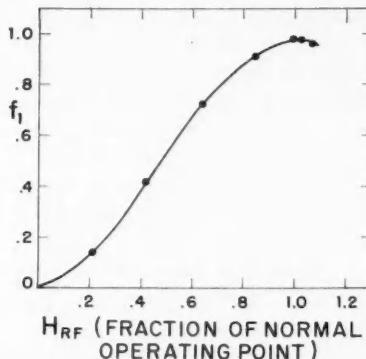


FIG. 6. The flipping probability  $f_1$  for first-order neutrons plotted against the r-f. current through the flipping coil.  $H_{RF}$  is the ratio of the current to the current for the optimum flipping probability.

the frequency of the r-f. was adjusted to give a minimum in  $r_1$ . The data of Fig. 6 confirm the sine-squared variation of flipping probability with r-f. current and show that at the optimum setting  $f_1 = 0.98 \pm 0.01$ .

The flipping probability  $f_2$  for the second-order neutrons was measured by setting the analyzing crystal to reflect such neutrons. With the current set to the optimum value for first-order flipping, the second-order flipping probability,  $f_2$ , was found to be  $.52 \pm .03$ . We would expect that  $f_2 = 0.5 f_1$  because the second-order velocity is twice the first-order velocity. We therefore prefer to assume that  $f_2 = .49 \pm .01$  and regard this measurement as a confirmation. The flipping probability  $f_3$  for the diffusely scattered neutrons was assumed to be 0.5.

The over-all flipping probability of the beam,  $\tilde{f}$ , can be related to the over-all polarization  $\tilde{P}$  in the unflipped condition by

$$\bar{P}(1-2\bar{f}) = \frac{a_1 P(1)(1-2f_1) + a_2 P(2)(1-2f_2) + a_3 P(3)(-2f_3)}{a_1 + a_2 + a_3}$$

where  $a_1$ ,  $a_2$ , and  $a_3$  are the numbers of first-order, second-order, and diffusely scattered neutrons. Using this formula the over-all beam flipping probability  $\bar{f}$  is  $0.97 \pm .03$ , where the error is mainly due to the uncertainty in the beam composition. The polarizations in the flipped conditions are listed in Table II.

### 5. MEASUREMENTS AWAY FROM THE BEAM CENTER

All the measurements of polarization and flipping probabilities discussed above were made with the analyzing crystal adjusted to look at the central part of the polarized beam. Measurements of the polarization with the r-f. off and of the flipping probability were repeated for first-order neutrons with the analyzer adjusted to look at the edge of the beam. The results were indistinguishable from those quoted above, indicating that the character of the beam is probably uniform over its cross section.

### 6. ACKNOWLEDGMENTS

We are very grateful to Dr. R. Nathans for many informative discussions during the early stages of the setting-up of this equipment. We are also very grateful to Mr. G. C. Dixon for his invaluable help in the fabrication and testing of the apparatus.

### REFERENCES

BACON, G. E. 1955. Neutron diffraction (Oxford University Press).  
 BLOCH, F. 1936. Phys. Rev. **50**, 529.  
 ——— 1937. Phys. Rev. **51**, 994.  
 BROCKHOUSE, B. N. 1959. Rev. Sci. Instr. **30**, 136.  
 BURGY, M. T., EPSTEIN, R. J., KROHN, V. E., NOVEY, T. B., RABOY, S., RINGO, G. R., and TELELDI, V. L. 1957. Phys. Rev. **107**, 1731.  
 BURGY, M. T., KROHN, V. E., NOVEY, T. B., RINGO, G. R., and TELELDI, V. L. 1958. Phys. Rev. **110**, 1214.  
 CLARK, M. A. and ROBSON, J. M. 1961. Can. J. Phys. **39**, 13.  
 CLARK, M. A., ROBSON, J. M., and NATHANS, R. 1958. Phys. Rev. Letters, **1**, 100.  
 CONNOR, D. 1959. Phys. Rev. Letters, **3**, 429.  
 DABBS, J. W. T., ROBERTS, L. D., and BERNSTEIN, S. 1955a. Oak Ridge Report ORNL 55-5-126.  
 ——— 1955b. Phys. Rev. **98**, 1512.  
 EKSTEIN, H. 1949. Phys. Rev. **76**, 1328.  
 HALPERN, O. 1949a. Phys. Rev. **75**, 343.  
 ——— 1949b. Phys. Rev. **76**, 1130.  
 HALPERN, O. and HOLSTEIN, T. 1941a. Phys. Rev. **59**, 960.  
 HALPERN, O., HAMERMESH, M., and JOHNSON, M. H. 1941b. Phys. Rev. **59**, 981.  
 HALPERN, O. and JOHNSON, M. H. 1939. Phys. Rev. **55**, 898.  
 HAMERMESH, M. 1942. Phys. Rev. **61**, 17.  
 ——— 1949. Phys. Rev. **75**, 1766.  
 HUGHES, D. J. and BURGY, M. T. 1951. Phys. Rev. **81**, 498.  
 HUGHES, D. J., WALLACE, J. R., and HOLTZMAN, R. H. 1948. Phys. Rev. **73**, 1277.  
 NATHANS, R. and PAOLETTI, 1959. Phys. Rev. Letters, **2**, 254.  
 NATHANS, R., SHULL, C. G., SHIRANE, G., and ANDRESON, A. 1959. J. Phys. Chem. Solids, **10**, 138.  
 SCHWINGER, J. 1937. Phys. Rev. **51**, 544.  
 SHULL, C. G. 1951. Phys. Rev. **81**, 626.  
 STANFORD, C. P., STEPHENSON, T. E., COCHRAN, L. W., and BERNSTEIN, S. 1954. Phys. Rev. **94**, 374.  
 STEINBERG, J. and WICK, G. C. 1949. Phys. Rev. **76**, 994.  
 STOLOVY, A. 1960. Phys. Rev. **118**, 211.  
 TAKEI, W. J., SHIRANE, G., and FRAZER, B. C. 1960. Phys. Rev. **119**, 122.  
 TRUMPY, G. 1957. Nuclear Phys., **2**, 644.  
 VERVERIER, J. 1960. Bull. Am. Phys. Soc. **5**, 239.

# THE ELECTRON ASYMMETRY IN THE BETA DECAY OF POLARIZED NEUTRONS<sup>1</sup>

M. A. CLARK AND J. M. ROBSON<sup>2</sup>

## ABSTRACT

The coefficient of the angular correlation between the electron direction and the neutron spin direction in the beta decay of the neutron has been measured using a beam of polarized neutrons. The coefficient of this correlation is  $-0.09 \pm 0.05$ . This implies that  $C_A/C_V$ , the ratio of the axial vector to vector coupling constants in neutron decay, is equal to  $-1.20 \pm 0.12$ .

## INTRODUCTION

In a recent note (Clark and Robson 1960) we described our measurements on the correlation between the neutron spin and the neutrino direction in the beta decay of the neutron and showed how this confirmed that the interaction was of the form  $(V-A)$  rather than  $(V+A)$ . In this paper we describe our measurements on the correlation between the neutron spin and the electron direction.

Jackson *et al.* (1957) have given theoretical expressions for the shape of the spectrum and for the angular correlations between the neutron decay products and the neutron spin. They show in particular that the correlation between the neutron spin and the electron direction should be of the form  $A \mathbf{J} \cdot \mathbf{p}_e / JE_e$  where  $\mathbf{J}$  is the spin of the neutron,  $\mathbf{p}_e$  and  $E_e$  are the momentum and energy of the electron, and  $A$  is a constant which depends on the coupling constants of the interactions.

If it is assumed that the decay proceeds by the axial vector and vector interactions only and that the two-component theory holds, then their expression for  $A$  becomes:

$$(1) \quad A = -2x(x+1)/(1+3x^2)$$

where  $x$  equals  $C_A/C_V$ , the ratio of the axial vector and vector coupling constants. Thus a measurement of  $A$  can give the ratio of the coupling constants directly without invoking any other experimental results. Measurements of  $A$  have been made by Burgy *et al.* (1957, 1958), whose latest value is  $-0.11 \pm 0.02$  (Telegdi 1958); this corresponds to  $x = -1.25 \pm 0.05$ .

## EXPERIMENTAL APPARATUS

Polarized neutrons were obtained by Bragg diffraction from a magnetized crystal of Co-Fe alloy mounted in a direct beam from the NRU reactor. The diffracted beam contained about  $3.5 \times 10^7$  neutrons per second with a polarization of about 93%. The direction of the neutron spins could be reversed by a

<sup>1</sup>Manuscript received September 28, 1960.

Contribution from the Physics Division, Atomic Energy of Canada Limited.

Issued as A.E.C.L. No. 1133.

<sup>2</sup>Present address: Physics Department, University of Ottawa, Ottawa, Ontario.

magnetic resonance technique without changing the steady state magnetic fields. A detailed description of the polarized beam is given in the preceding paper (Clark and Robson 1961).

A vertical section through the apparatus used in this experiment is shown in Fig. 1; its location along the beam of polarized neutrons can be seen from Fig. 1 of the preceding paper. In Fig. 1 of this paper the neutron beam is

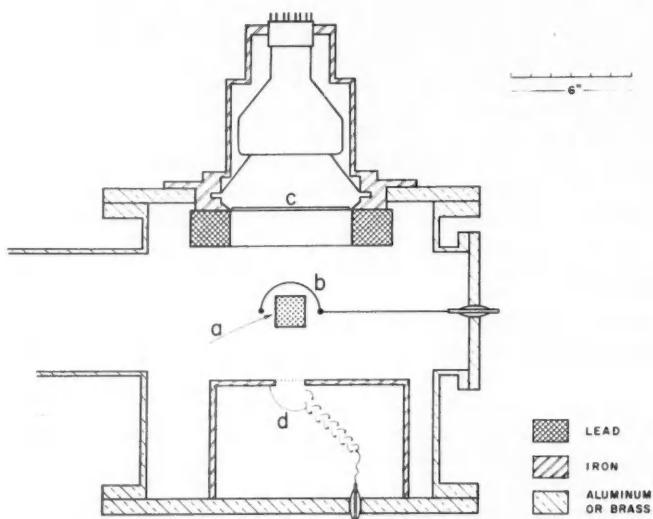


FIG. 1. Vertical cross section through the experimental apparatus. (a) The beam of polarized neutrons, shown in cross section, should be visualized as coming up out of the paper. (b) High voltage focussing electrode. (c) Electron detector. (d) Proton detector. The large pipe at the left leads to conventional high vacuum apparatus.

shown in cross section and is about  $1\frac{1}{2}$  in. square. The spins are 'down' if the radio-frequency coil described in the preceding paper is not excited, and are 'up' if it is excited. The electrons from the decay of these neutrons are detected by a 6-in. diameter plastic scintillation counter mounted above the beam and clearly visible in Fig. 1. In principle the coefficient  $A$  could be measured by comparing the counting rate of this counter with the neutron spins 'up' to that with the spins 'down'. However, the large background counting rate necessitates the use of a coincidence technique to separate out those events which are due to neutron decays; therefore the low energy recoil protons resulting from the decays are electrostatically focussed from the beam onto the first dynode of an electron multiplier and are used to identify the corresponding pulses from the electron counter. This identification is accomplished by a delayed-coincidence technique in which events are counted only if the pulse from the electron multiplier is delayed from the scintillation-counter pulse by the correct transit time for a proton through the electrostatic

collecting field. The details of these parts of the apparatus will now be more fully described.

The scintillation counter used a NE102\* plastic phosphor  $\frac{1}{8}$  in. thick and 6 in. in diameter. It was mounted without oil on a light pipe made of non-ultra-violet-absorbing plexiglas whose shape can be visualized from Fig. 1. The light pipe was in turn mounted on a DuMont 5-in. photomultiplier tube type 6364; oil was used as an optical coupling between the light pipe and the photomultiplier. The dry mounting of the scintillator sheet caused an appreciable amount of light to be guided by total internal reflection to the circular edge. Here the sloping sides of the light pipe reflected this light to the photocathode of the photomultiplier tube. This effect served to boost the light collection from points near the circumference of the scintillator, and hence to make light collection more uniform. The differential pulse height distribution obtained with a source of  $\text{Sn}^{113}$  is shown in Fig. 2 and indicates a resolution of about 43% for the 364-kev conversion electrons.

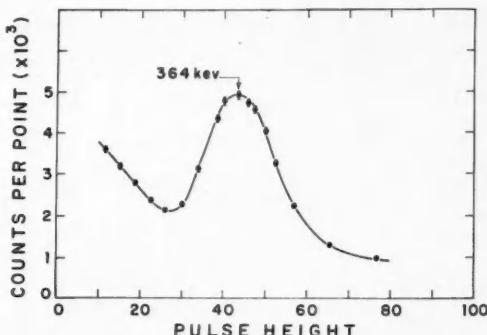


FIG. 2. The differential pulse height distribution obtained from the 6-in. diameter plastic scintillation counter using a source of  $\text{Sn}^{113}$  mounted at the center of the neutron decay region.

The proton focussing system consisted of a high voltage electrode in the shape of a half cylinder 3 in. in diameter and 6 in. in length which is shown in section (b) in Fig. 1. This electrode, made of .0004-in. aluminized mylar held at +10 kv, was essentially around the beam and directly opposite the  $1\frac{1}{2}$  in. by 1 $\frac{1}{2}$  in. entrance aperture to the electron multiplier. Tests made by rolling steel balls on a rubber-sheet model indicated that all protons emitted by neutrons decaying in a 2-cm length of beam would be focussed through the entrance aperture onto the first dynode of the electron multiplier.

Delayed coincidences between the proton and electron counters were recorded by a time sorter using a conventional 'fast-slow' system. The 'fast' pulses were fed directly to a time to pulse height converter which in turn fed a 100-channel pulse height analyzer. The analyzer was gated by the 'slow' pulses from a single channel analyzer on the electron counter and from a

\*Nuclear Enterprises Limited, Winnipeg, Manitoba.

discriminator on the proton counter. Thus a 'window' could be set to record electrons having energies between two given limits corresponding to the required part of the neutron decay spectrum, and the proton discriminator could reject the small pulses due to electrons and gamma rays striking the proton counter. An anticoincidence circuit was also included to enable the 100-channel pulse height analyzer to reject events associated with many large pulses from the time to pulse height converter which corresponded to proton counter pulses unaccompanied by a corresponding pulse from the electron counter arriving within the active time of the converter. These, in random coincidence with randomly generated gating pulses from the slow system, would give a large contribution to the flat background of random events appearing in the time sorter spectrum unless eliminated by the anticoincidence. A block diagram of the coincidence system is shown in Fig. 3. Individual scalers (not shown in Fig. 3) were used to record the accumulated counts from each detector, from a fission neutron detector mounted in the polarized neutron beam and from the boron trifluoride neutron counter used after the polarization analyzer.

The contents of all data scalers and the 100-channel pulse height analyzer were read out on a 7-hole punched tape. The data were then transferred directly

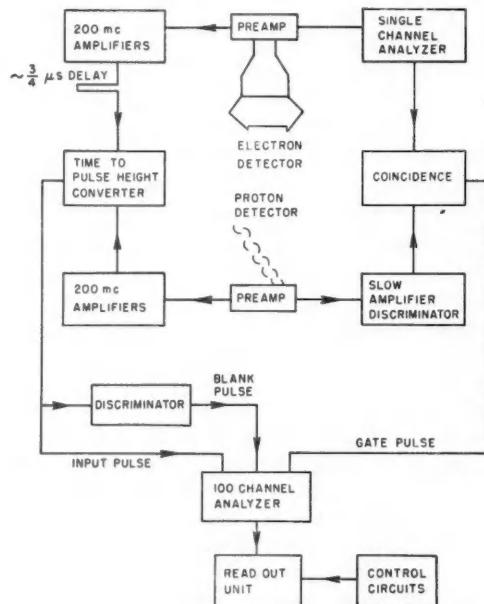


FIG. 3. Block diagram of the 'fast-slow' time sorting unit. The scalers and monitor recorders have been omitted from the diagram for simplicity.

into the Chalk River datatron computer where routine summing operations were performed to produce results corresponding to neutron spins 'up' and 'down'.

The sequence of the experiment was to accumulate data for 1 hour with the neutron spins 'up', read out this data onto punched tape, and then repeat for 1 hour with the spins 'down'. The control system for these operations was activated by an electric clock and after 1 hour of reactor operation it automatically went through the sequence of stopping further data accumulation, printing out all data, resetting, activating the oscillator which reversed the neutron spins, and restarting the count.

The vacuum system was conventional and maintained the pressure in the apparatus at less than  $1 \times 10^{-6}$  mm Hg throughout the experiment. A system of safety protection devices enabled the apparatus to be left unattended and in full operation overnight, thus permitting almost continuous accumulation of data.

#### RESULTS

Figure 4 shows the accumulated data for a series of runs taken with the window of the electron-counter pulse height analyzer set to accept electrons with energies between 200 kev and 700 kev. The bottom curve shows data

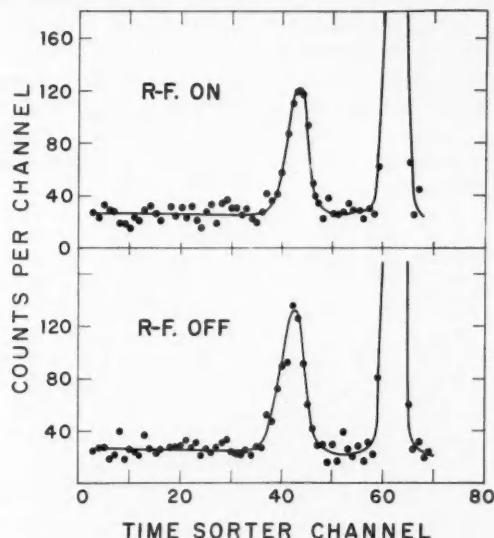


FIG. 4. Coincidences between the proton counter and electron counter plotted as a time spectrum. The top curve is for neutrons whose spins are 'up' and the bottom curve is for 'down' neutrons. The large peaks at about channel 62 are prompt coincidence peaks due to scattering events in the vacuum chamber and the wider peaks around channel 42 are due to neutron decay events. One channel of the time sorter corresponds to about 9  $\mu$ sec and events in which the proton counter pulse occur later than the electron counter pulse appear to the left of the prompt peaks.

taken with the neutron spins 'down' and the top curve shows data taken with the spins 'up'. There is a large prompt coincidence peak at channel 62 which is probably due to Compton scattering events recorded in both counters. This peak appears whenever the detectors are exposed to a high flux of energetic gamma rays, even in the absence of the neutron beam. Motion to the left of the prompt peak on these graphs corresponds to delay of the proton detector pulse relative to the electron detector pulse and the peak near channel 42 is delayed by about 180  $\mu$ sec which is appropriate for the transit time of the protons resulting from the neutron decay. This peak moved in position as the voltage on the proton focussing electrode was changed, depended on the presence of neutrons, and is undoubtedly due to coincidences between the protons and electrons from the beta decay of polarized neutrons. Data were also accumulated with the electron-counter window set to accept electrons between 300 kev and 700 kev.

The data were used to compute a coincidence asymmetry,  $a$ , which is defined as:

$$(2) \quad a = [C(\text{off}) - C(\text{on})]/[C(\text{off}) + C(\text{on})]$$

where  $C(\text{off})$  and  $C(\text{on})$  are the number of coincidences with the neutron spins 'down' and 'up', normalized to the same number of neutrons passing through the apparatus. This normalization was made on the basis of the fission monitor counts. The numbers of coincidences were obtained from the areas under the neutron peaks using the flat regions to the left of the neutron peaks and between the neutron and prompt peaks to obtain estimates of the random coincidence background.

The coincidence asymmetry,  $a$ , is related to the  $A$  coefficient of equation (1) by the following formula:

$$(3) \quad a = \frac{\{P(\text{off}) - P(\text{on})\} A \left( \frac{v}{c} \cos \theta \right)}{2 + \{P(\text{off}) + P(\text{on})\} A \left( \frac{v}{c} \cos \theta \right)}.$$

In this equation  $v$  is the velocity of an electron which makes an angle  $\theta$  with the upwards vertical, and the average is taken over all possible cases within the geometric limitations of the apparatus and within the window of the electron counter channel.  $P(\text{off})$  and  $P(\text{on})$  are the neutron beam polarizations with the radio-frequency flipping coil unexcited and excited. Since  $P(\text{off}) \approx -P(\text{on})$  and  $A$  turns out to be small equation (3) can be approximated by:

$$(4) \quad P.A. \left( \frac{v}{c} \cos \theta \right)$$

where  $P$  is the average of the polarizations in the two beam conditions: neutron polarization 'up' and 'down'.  $(v/c) \cos \theta$  was computed from the known shape of the spectrum and the known geometry for use in equation (4).

Table I gives the measured values of  $a$ , the computed values of the

TABLE I  
Experimental values of the electron asymmetry

Electron energy range (kev)	$\left(\frac{\bar{v}}{c}\right)$	$\langle \cos \theta \rangle$	$P$	$a$	$A$
200 → 700	$0.80 \pm .03$	$0.91 \pm .04$	$-.89 \pm .05$	$+.050 \pm .044$	$-.076 \pm .068$
300 → 700	$0.83 \pm .03$	$0.91 \pm .04$	$-.89 \pm .05$	$+.070 \pm .039$	$-.103 \pm .057$

coefficients in equation (4), and the values of  $A$  derived from this equation. The final average for  $A$  is  $-0.09 \pm .05$ . The value of this uncertainty, which can be interpreted as a standard deviation, includes allowances for all known experimental errors except the collection efficiency of the protons. We have assumed that this efficiency is independent of all the proton recoil angles involved in the detection of the electron proton coincidences and base this assumption on our experience with the rolling of steel balls on a rubber model of the electrostatic system. The agreement between the two runs covering different electron energy ranges is a reassuring confirmation of this since the range of the proton-electron angles will be much greater for the first run than for the second. Also some preliminary runs with different potentials on the high voltage focussing electrode showed no strong dependence of the recorded number of decay events with this focussing potential again suggesting complete collection efficiency. However, the work of Burgy *et al.* (1957, 1958) shows how significant this type of error can be and we feel that the proton collection uncertainty restricts the ultimate accuracy of this type of experiment to about the error quoted in Table I.

#### DISCUSSION

The experimental result of  $A = -0.09 \pm 0.05$  can be used to compute  $x = C_A/C_V$  from equation (1). The possible values of  $x$  are  $-1.20 \pm 0.12$  and  $+0.05 \pm 0.05$ . The value of  $x^2$  derived by Kistner and Rustad (1959) from the neutron half-life measurement of Sosnovsky *et al.* (1959) is  $1.42 \pm 0.06$  and therefore rules out the second solution of equation (1) and restricts the value of  $x$  to  $-1.20 \pm 0.12$ . This value and the equation (1) are shown in Fig. 5, which is a plot of  $A$  against  $x$  for the neutron decay assuming that it proceeds only by a mixture of the axial vector and vector interactions. The value of  $|x| = |C_A/C_V|$  derived from the neutron half life is also shown.

There are now several different measurements of the various neutron decay parameters which can be used to derive values of  $x = C_A/C_V$ ; these are listed in Table II together with the derived values of  $x$  or  $|x|$  assuming A and V interactions only. The neutrino asymmetry coefficient  $B$  is very insensitive to  $x$  in the region of interest (Clark and Robson 1960), and is merely included for completeness. It can only be used to give a center value and an upper limit on  $x$ ; these are indicated in Table II.

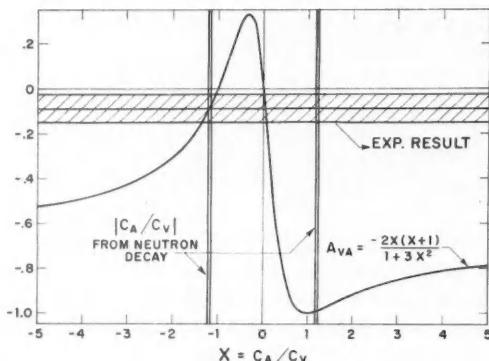


FIG. 5. The solid line shows  $A$  plotted against  $x$  for the neutron decay assuming only axial vector and vector interactions. The shaded horizontal region is the value of  $A$  obtained in this experiment and the two narrow vertical unshaded zones indicate the value of  $|x|$  obtained from the half life of the neutron (Kistner and Rustad 1959; Sosnovsky *et al.* 1959).

TABLE II  
Comparison of neutron decay experiments from which  $x$  may be derived

Quantity measured	Value	$ x $	$x$	Ref.
Half life	$11.7 \pm 0.3$ m	$1.19 \pm .03$		<i>a, b</i>
Angular correlation	$+0.07 \pm 0.12$	$0.88 \pm 0.24$ $0.16$		<i>c</i>
"	$-0.06 \pm 0.13$	$1.13 \pm 0.53$ $0.25$		<i>d</i>
Electron asymmetry	$-0.11 \pm 0.02$ $-0.09 \pm 0.05$		$-1.25 \pm .05$ $-1.20 \pm .12$	<i>e</i> <i>f</i>
Neutrino asymmetry	$0.88 \pm 0.15$ $+0.96 \pm 0.40$		$-2.6 ; < -0.4$ $-1.6 ; < -0.3$	<i>g</i> <i>h</i>

*a.* Sosnovsky *et al.* (1959).  
*b.* Kistner and Rustad (1959).  
*c.* Robson (1958).  
*d.* Trebukhovskii *et al.* (1959).  
*e.* Teleghi (1958) and Burgy *et al.* (1958).  
*f.* This work.  
*g.* Burgy *et al.* (1958).  
*h.* Clark and Robson (1960).

#### ACKNOWLEDGMENTS

We are grateful to Mr. W. D. Howell for designing and setting up the automatic data read-out system, to Dr. J. M. Kennedy and his computer group for assisting with the data analysis, and to Mr. G. C. Dixon for his assistance in testing and using the experimental apparatus.

## REFERENCES

BURGY, M. T., EPSTEIN, R. J., KROHN, V. E., NOVEY, T. B., RABOY, S., RINGO, G. R., and TELELDI, V. L. 1957. *Phys. Rev.* **107**, 1731.

BURGY, M. T., KROHN, V. E., NOVEY, T. B., RINGO, G. R., and TELELDI, V. L. 1958. *Phys. Rev.* **110**, 1214.

CLARK, M. A. and ROBSON, J. M. 1960. *Can. J. Phys.* **38**, 693.

— 1961. *Can. J. Phys.* **39**, 1.

JACKSON, J. D., TREIMAN, S. B., and WYLD, H. W., JR. 1957. *Phys. Rev.* **106**, 517.

KISTNER, O. C. and RUSTAD, B. M. 1959. *Phys. Rev.* **114**, 1329.

ROBSON, J. M. 1958. *Can. J. Phys.* **36**, 1450.

SOSNOVSKY, A. N., SPIVAK, P. E., PROKOFIEV, YU. A., KUTIKOV, I. E., and DOBRININ, YU. P. 1959. *Nuclear Physics*, **10**, 395.

TELEGDI, V. L. 1958. Conference on Weak Interactions, Gatlinbury Paper D. I.

TREBUKHOVSKII, YU. V., VLADIMIVSKII, V. V., GRIGORIEV, V. K., and ERGAKOV, V. A. 1959. *J. Exptl. Theoret. Phys. (U. S. S. R.)* **36**, 1314.

# REPRESENTATION OF SPACE INVERSION, TIME REVERSAL, AND PARTICLE CONJUGATION IN QUANTUM FIELD THEORY<sup>1</sup>

F. A. KAEMPFER

## ABSTRACT

The unitary operators of space inversion and particle conjugation and the unitary factor of the antiunitary operator of time reversal can each be written in the form  $e^{i\Omega}$ , where  $\Omega$  is the direct sum of two terms,  $\Omega = \Omega_1\theta_1 + \Omega_2\theta_2$ , with  $\Omega_1, \Omega_2$  Hermitean bilinear forms in the creation and annihilation operators of the boson or fermion field under consideration, and  $\theta_1, \theta_2$  singular operators which separate the appropriate half spaces needed for the formulation of the symmetry operations. Explicit expressions are given for the generators  $\Omega$  in case of a non-Hermitean boson field of spin 0, and in case of a four-component fermion field of spin  $\frac{1}{2}$ .

## 1. INTRODUCTION

An examination of current literature on quantum field theory (see, for example, Jauch and Rohrlich 1955; Källen 1958; Bogoliubov and Shirkov 1959; Hamilton 1959; Mandl 1959) will reveal a startling dichotomy in the representation of observables.

All authors write down representations bilinear in annihilation and creation operators for operators identified with mechanical observables, energy, momentum, angular momentum, thus giving substance to the notion of particles which are thought of as carriers of the mechanical properties of the fields considered. Such explicit representations are treated and employed at least as extensively as are the symmetry properties which allow one to write formally these observables as single symbols  $P_\mu, S_{ik}$ , standing for generators of unitary transformations representing displacements and rotations in space and time.

This contrasts with the treatment accorded those operators which are associated with the observable parity, reversality, and conjugality properties of the fields. They are defined and employed almost exclusively as symbolic generators of the unitary transformations of space inversion and particle conjugation and of the antiunitary transformation of time reversal, without being broken down further and represented in terms of the appropriate annihilation and creation operators. Although attempts have been made to produce representations of these operators in terms of creation and annihilation operators by Watanabe (1955) and in case of particle conjugation already by Wolfenstein and Ravenhall (1952),\* the work of these authors remained incomplete and their results have either been ignored altogether or relegated to a footnote (Hamilton 1959, p. 178).

The purpose of the work reported in this paper was to find explicit expressions, valid in the entire space spanned by the state vectors, for the operators of space inversion  $\Pi$ , time reversal  $\Theta$ , and particle conjugation  $\Gamma$ , by writing them as functions of bilinear forms in the appropriate annihilation and creation

<sup>1</sup>Manuscript received August 3, 1960.

Contribution from the Department of Physics, University of British Columbia, Vancouver 8, B.C.

\*See also Nigam and Foldy (1956).

operators, thus amending and supplementing the work of Watanabe (1955), whose expressions are valid in certain half spaces only.

The advantage of having such representations available is obvious when one wants to construct many-particle states  $|\rangle$  of specific parity  $P$ , namely the solutions of the eigenvalue problem

$$(1.1) \quad \Pi|\rangle = P|\rangle,$$

states of specific reversality  $T$ , defined by the equation

$$(1.2) \quad \Theta|\rangle = \langle |T,$$

and states of specific conjugality  $C$ , satisfying

$$(1.3) \quad \Gamma|\rangle = C|\rangle.$$

Conversely, for a given many-particle state  $|\rangle$  the representations of  $\Pi$ ,  $\Theta$ ,  $\Gamma$  in terms of particle annihilation and creation operators enable one to find the parity, reversality, and conjugality properties of that state by straightforward computation.

Many selection rules governing interacting systems follow from the commutation relations (C.R.'s) of the operators  $\Pi$ ,  $\Theta$ ,  $\Gamma$  with operators representing mechanical and other observables. The existence of representations for the latter in terms of annihilation and creation operators is of no use for the computation of these C.R.'s, unless such representations are also available for  $\Pi$ ,  $\Theta$ ,  $\Gamma$ .

Finally, problems arising from the so-called TCP-theorem (Lüders 1954) require knowledge of the C.R.'s of  $\Pi$ ,  $\Theta$ ,  $\Gamma$  among themselves, and again availability of explicit representations for these operators may be useful as a computational aid.

## 2. SUMMARY OF FORMALISM

Calculations will be confined to bosons of spin 0 and mass  $m$  described by a one-component field operator  $\phi(x)$ , and to fermions of spin  $\frac{1}{2}$  and mass  $M$  described by a four-component field operator  $\psi(x)$ . Extension of the results to fields of higher spin, laborious though it may be, does not encounter any difficulties beyond those found in the treatment of these two basic cases. The boson field  $\phi(x)$  is not assumed to be Hermitean, it includes thus the case of neutral particles of spin 0 which differ from their antiparticles. The well-known (Källen 1958) invariant decomposition into positive and negative frequency parts belonging to the various modes of propagation  $k$  will be employed.

The boson field  $\phi(x)$  is written as a sum of contributions, normalized in a unit volume,

$$(2.1) \quad \phi(x) = \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega}} \left\{ b_{\mathbf{k}}(1)e^{ikx} + b_{\mathbf{k}}^+(2)e^{-ikx} \right\}$$

where

$$(2.2) \quad kx = \mathbf{k}x - \omega t$$

and

$$(2.3) \quad \omega = +\sqrt{(\mathbf{k}^2 + m^2)}.$$

The symbol  $b_{\mathbf{k}}(r)$  represents the annihilation operator of a particle ( $r = 1$ ) or antiparticle ( $r = 2$ ) of momentum  $\mathbf{k}$ , and its Hermitean conjugate  $b_{\mathbf{k}}^{\dagger}(r)$  represents the corresponding creation operator. The Hermitean conjugate  $\phi^+(x)$  of the field operator (2.1) is, in general, different from  $\phi(x)$ , namely

$$(2.4) \quad \phi^+(x) = \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega}} \left\{ b_{\mathbf{k}}(2)e^{ikx} + b_{\mathbf{k}}^{\dagger}(1)e^{-ikx} \right\}.$$

The C.R.'s governing the annihilation and creation operators are

$$(2.5) \quad [b_{\mathbf{k}}(r), b_{\mathbf{k}'}^{\dagger}(r')]_{-} = \delta_{\mathbf{kk}'}\delta_{rr'};$$

all other commutators = 0, giving rise to a representation of boson states  $|n_{\mathbf{k}}(r)\rangle$  with  $n_{\mathbf{k}}(r)$  particles ( $r = 1$ ) or antiparticles ( $r = 2$ ) of momentum  $\mathbf{k}$  present, in which these operators have the properties

$$(2.6) \quad b_{\mathbf{k}}(r)|n_{\mathbf{k}}(r)\rangle = \sqrt{n_{\mathbf{k}}(r)}|n_{\mathbf{k}}(r)-1\rangle; \quad b_{\mathbf{k}}^{\dagger}(r)|n_{\mathbf{k}}(r)\rangle = \sqrt{n_{\mathbf{k}}(r)+1}|n_{\mathbf{k}}(r)+1\rangle.$$

A state containing a given set of particles and antiparticles of various momenta is the direct product

$$(2.7) \quad |\dots n \dots\rangle = \prod_{\mathbf{k}} \prod_{r=1}^2 |n_{\mathbf{k}}(r)\rangle,$$

so that the most general boson state can be written as a linear superposition

$$(2.8) \quad |\rangle = \sum_{\dots n \dots} B(\dots n \dots) |\dots n \dots\rangle$$

where the summation goes over all possible sets of occupation numbers  $n$ . The vacuum state is denoted by  $|0\rangle$ , from which one can form, in accordance with (2.6), by application of the appropriate creation operators, any many-particle state (2.7).

The spinor field  $\psi(x)$  is similarly decomposed and written, normalized in a unit volume,

$$(2.9) \quad \psi(x) = \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega}} \sum_{s=1}^2 \left\{ a_{\mathbf{k}}(s,1)u_{\mathbf{k}}(s,1)e^{ikx} + a_{\mathbf{k}}^{\dagger}(s,2)u_{\mathbf{k}}(s,2)e^{-ikx} \right\}$$

where  $kx$  is defined as in equation (2.2)

$$(2.10) \quad \omega = +\sqrt{(\mathbf{k}^2 + M^2)}$$

and the four four-component spinor amplitudes  $u_{\mathbf{k}}(s,r)$  are given by

$$(2.11) \quad u_{\mathbf{k}}(1,1) = \begin{pmatrix} 1 \\ 0 \\ k_3 \\ k_1 + ik_2 \end{pmatrix} \quad u_{\mathbf{k}}(2,1) = \begin{pmatrix} 0 \\ 1 \\ k_1 - ik_2 \\ -k_3 \end{pmatrix} \quad u_{\mathbf{k}}(1,2) = \begin{pmatrix} k_3 \\ k_1 + ik_2 \\ 1 \\ 0 \end{pmatrix}$$

$$u_{\mathbf{k}}(2,2) = \begin{pmatrix} k_1 - ik_2 \\ -k_3 \\ 0 \\ 1 \end{pmatrix}.$$

With these conventions of phases and labels the symbol  $a_{\mathbf{k}}(s,r)$  represents the annihilation operator of a particle ( $r = 1$ ) or antiparticle ( $r = 2$ ) of momentum  $\mathbf{k}$  with spin in  $k_3$ -direction ( $s = 1$ ) or in  $-k_3$ -direction ( $s = 2$ ), and  $a_{\mathbf{k}}^+(s,r)$  represents the corresponding creation operator. The adjoint operator  $\bar{\psi} = \psi^+ \gamma_4$  will be written

$$(2.12) \quad \bar{\psi}(x) = \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega}} \sum_{s=1}^2 \left\{ a_{\mathbf{k}}^+(s,1) \bar{u}_{\mathbf{k}}(s,1) e^{-ikx} + a_{\mathbf{k}}(s,2) \bar{u}_{\mathbf{k}}(s,2) e^{ikx} \right\}$$

where the adjoint spinor amplitudes  $\bar{u} = u^* \gamma_4$  are in components

$$(2.13) \quad \begin{aligned} \bar{u}_{\mathbf{k}}(1,1) &= \begin{pmatrix} 1 \\ 0 \\ -k_3 \\ -k_1 + ik_2 \end{pmatrix} & \bar{u}_{\mathbf{k}}(2,1) &= \begin{pmatrix} 0 \\ 1 \\ -k_1 - ik_2 \\ k_3 \end{pmatrix} & \bar{u}_{\mathbf{k}}(1,2) &= \begin{pmatrix} k_3 \\ k_1 - ik_2 \\ -1 \\ 0 \end{pmatrix} \\ & & & & & \\ \bar{u}_{\mathbf{k}}(2,2) &= \begin{pmatrix} k_1 + ik_2 \\ -k_3 \\ 0 \\ -1 \end{pmatrix}. \end{aligned}$$

The operations of space inversion, time reversal, and particle conjugation, as far as the spinor components are concerned, involve the matrices

$$(2.14) \quad \Lambda_{\Pi} = \gamma_4 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

$$(2.15) \quad \Lambda_{\Theta} = \gamma_1 \gamma_3 = -i\sigma_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$(2.16) \quad \Lambda_{\Gamma} = \gamma_2 \gamma_4 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

which have the following important properties

$$(2.17) \quad \begin{cases} \Lambda_{\Pi} u_{\mathbf{k}}(s,1) = u_{-\mathbf{k}}(s,1) \\ \Lambda_{\Pi} u_{\mathbf{k}}(s,2) = -u_{-\mathbf{k}}(s,2) \end{cases}$$

$$(2.18) \quad \begin{cases} \Lambda_{\Theta} u_{\mathbf{k}}^*(1,r) = u_{-\mathbf{k}}(2,r) \\ \Lambda_{\Theta} u_{\mathbf{k}}^*(2,r) = -u_{-\mathbf{k}}(1,r) \end{cases}$$

$$(2.19) \quad \begin{cases} \Lambda_{\Gamma} \bar{u}_{\mathbf{k}}(s,r) = u_{\mathbf{k}}(s',r') \\ \Lambda_{\Gamma} \bar{u}_{\mathbf{k}}(s,r) = -u_{\mathbf{k}}(s',r') \end{cases} \quad \begin{array}{l} \text{for } s \neq r \\ \text{for } s = r \end{array} \quad \begin{cases} s' \neq s; r' \neq r. \end{cases}$$

The relations (2.17)–(2.19) are invariant under unitary transformations in spinor space.

In accordance with the exclusion principle the annihilation and creation operators satisfy the anti C.R.'s

$$(2.20) \quad \{a_{\mathbf{k}}(s,r), a_{\mathbf{k}'}^+(s',r')\}_+ = \delta_{\mathbf{kk}'}\delta_{ss'}\delta_{rr'},$$

all other anticommutators = 0, giving rise to a representation of fermion states  $|N_{\mathbf{k}}(s,r)\rangle$  with  $N_{\mathbf{k}}(s,r)$  particles ( $r = 1$ ) or antiparticles ( $r = 2$ ) of momentum  $\mathbf{k}$  and spin label  $s$  present, the occupation number  $N$  being restricted to the values 0 and 1. In this representation the operators have the properties

$$(2.21) \quad a_{\mathbf{k}}(s,r)|1_{\mathbf{k}}(s,r)\rangle = c_{\mathbf{k}}(s,r)|0_{\mathbf{k}}(s,r)\rangle; \quad a_{\mathbf{k}}^+(s,r)|0_{\mathbf{k}}(s,r)\rangle = c_{\mathbf{k}}^*(s,r)|1_{\mathbf{k}}(s,r)\rangle$$

where  $c_{\mathbf{k}}(s,r)$  is the well-known sign factor. A state containing a given set of particles and antiparticles of various momenta and spins is the direct product

$$(2.22) \quad |\dots N \dots\rangle = \prod_{\mathbf{k}} \prod_{s=1}^2 \prod_{r=1}^2 |N_{\mathbf{k}}(s,r)\rangle.$$

Unoccupied states are included in this product and the ordering convention imposed by the sign factor must be maintained through any calculation. The most general fermion state can then be written, in analogy to (2.8)

$$(2.23) \quad |\rangle = \sum_{\dots N \dots} A(\dots N \dots) |\dots N \dots\rangle$$

where the summation goes over all possible sets of occupation numbers  $N$ , including all sets in which any number of  $N$ 's may be zero.

### 3. REPRESENTATION OF $\Pi$ , $\Theta$ , $\Gamma$ FOR A BOSON FIELD

The operators of space inversion, time reversal, and particle conjugation for the spinless boson field (2.1) are defined by the equations (see, for example, Corinaldesi 1958)

$$(3.1) \quad \Pi\phi(\mathbf{x},t)\Pi^{-1} = \eta_{\Pi}\phi(-\mathbf{x},t)$$

$$(3.2) \quad \Theta\phi(\mathbf{x},t)\Theta^{-1} = \eta_{\Theta}^*\phi^+(\mathbf{x},-t)$$

$$(3.3) \quad \Gamma\phi(\mathbf{x},t)\Gamma^{-1} = \eta_{\Gamma}^*\phi^+(\mathbf{x},t).$$

From the assumption that each of these operators applied twice should result in the identity operation follow conditions on the factors  $\eta$ , namely

$$(3.4) \quad \eta_{\Pi}^2 = 1$$

$$(3.5) \quad \eta_{\Theta}\eta_{\Theta}^* = 1$$

$$(3.6) \quad \eta_{\Gamma}\eta_{\Gamma}^* = 1.$$

The operators  $\Pi$ ,  $\Theta$ ,  $\Gamma$  act, by definition, solely on the  $b_{\mathbf{k}}$  and  $b_{\mathbf{k}}^+$  and do not affect any  $c$ -numbers in  $\phi$ , thus, for example,  $\Theta(b e^{ikx})\Theta^{-1} = \Theta b \Theta^{-1} e^{ikx}$  etc. The antiunitary operator  $\Theta$  may be decomposed, following Schwinger (1951), into a unitary operator  $U$  and the nonlinear operator  $K$  of transposition,

$$(3.7) \quad \Theta = UK.$$

$K$  changes kets into bras and vice versa, and reverses the order of all operator factors. Thus, if  $b_1$  and  $b_2$  are two operators,

$$(3.8) \quad Kb_1b_2| = \langle \tilde{b}_2\tilde{b}_1$$

where  $\tilde{b}$  denotes the transpose of the operator  $b$ . It is, perhaps, instructive to verify that this definition agrees with the one given by Lüders (1957) and others, who define time reversal as

$$(3.9) \quad T\phi(\mathbf{x},t)T^{-1} = \eta_\Theta\phi(\mathbf{x},-t)$$

where  $T$  is the antiunitary operator made up out of the unitary operator  $U$  and the nonlinear operator  $L$  of complex conjugation. Indeed, by substituting (3.7) into (3.2) one obtains, using  $\phi^+ = \tilde{\phi}^*$  and  $U^{-1} = U^+$ ,

$$(3.10) \quad \begin{aligned} UK\phi(\mathbf{x},t)K^{-1}U^{-1} &= U\tilde{\phi}(\mathbf{x},t)U^{-1} = (U\phi^*(\mathbf{x},t)U^{-1})^+ \\ &= (UL\phi(\mathbf{x},t)L^{-1}U^{-1})^+ = (T\phi(\mathbf{x},t)T^{-1})^+ = \eta_\Theta^*\phi^+(\mathbf{x},-t) \end{aligned}$$

which is the adjoint of equation (3.9).

Writing down equations (3.1), (3.2), (3.3) in terms of the expansions (2.1) and (2.4), changing the summation index on the right-hand sides of (3.1) and (3.2) from  $\mathbf{k}$  to  $-\mathbf{k}$ , one can compare coefficients and obtain the following relations which give the effect of  $\Pi$ ,  $\Theta$ ,  $\Gamma$  on the annihilation and creation operators:

$$(3.11) \quad \Pi b_{\mathbf{k}}(1)\Pi^{-1} = \eta_\Pi b_{-\mathbf{k}}(1); \quad \Pi b_{\mathbf{k}}(2)\Pi^{-1} = \eta_\Pi^* b_{-\mathbf{k}}(2)$$

$$(3.12) \quad \Theta b_{\mathbf{k}}(1)\Theta^{-1} = \eta_\Theta^* b_{-\mathbf{k}}^+(1); \quad \Theta b_{\mathbf{k}}(2)\Theta^{-1} = \eta_\Theta b_{-\mathbf{k}}^+(2)$$

$$(3.13) \quad \Gamma b_{\mathbf{k}}(1)\Gamma^{-1} = \eta_\Gamma b_{\mathbf{k}}(2); \quad \Gamma b_{\mathbf{k}}(2)\Gamma^{-1} = \eta_\Gamma b_{\mathbf{k}}(1).$$

Equations (3.12) contain the effect of  $U$  on the  $b_{\mathbf{k}}$ . Using (3.7) and the fact that, because of the reality of the  $b_{\mathbf{k}}$ ,  $\tilde{b}_{\mathbf{k}} = b_{\mathbf{k}}^+$ , one obtains

$$(3.14) \quad Ub_{\mathbf{k}}(1)U^{-1} = \eta_\Theta b_{-\mathbf{k}}(1); \quad Ub_{\mathbf{k}}(2)U^{-1} = \eta_\Theta^* b_{-\mathbf{k}}(2),$$

which are seen to be identical in structure with the equations (3.11) for  $\Pi$ . It should be stressed, however, that  $\Pi$  and  $U$  are not identical. The essential difference between  $\eta_\Pi$  and  $\eta_\Theta$ , contained already in equations (3.4) and (3.5), arises from the condition  $\Pi^2 = 1$  on one hand, imposed on  $\Pi$  in addition to unitarity  $\Pi^{-1} = \Pi^+$ , making it Hermitean  $\Pi = \Pi^{-1} = \Pi^+$  and thus requiring  $\eta_\Pi$  to be real, whereas the condition  $\Theta^2 = 1$ , on the other hand, leads only to the relation  $UU^* = 1$ , so that  $\eta_\Theta$  need not be real. Indeed, the operation  $\Theta^2\phi\Theta^{-2}$  reads explicitly

$$(3.15) \quad UK(UK\phi K^{-1}U^{-1})K^{-1}U^{-1} = UK(U\tilde{\phi}U^{-1})K^{-1}U^{-1} = \widetilde{U\widetilde{U^{-1}}\phi\widetilde{U}U^{-1}}$$

which reduces to the identity only if

$$(3.16) \quad \widetilde{U\widetilde{U^{-1}}} = \widetilde{U\widetilde{U^+}} = UU^* = 1.$$

Quite generally, the antiunitary character of  $\Theta$  requires  $U$  to transform under any unitary transformation  $S$  of the state vectors as (see Jauch and Rohrlich 1955, p. 92)

$$(3.17) \quad U' = S^* U S^{-1}.$$

If, in particular,  $S = U$ , one obtains  $U' = U^*$ . One may thus say that if the transformation  $t \rightarrow -t$  is associated with  $U$ , then the transformation  $-t \rightarrow t$  is associated with  $U^*$ .

Construction of representations for  $\Pi$ ,  $U$ ,  $\Gamma$  is facilitated if these unitary operators are written in the form  $e^{i\alpha\Omega}$ , where the generators  $\Omega$  are Hermitean operators which will be labelled  $\Omega_{\Pi}$ ,  $\Omega_{\Theta}$ ,  $\Omega_{\Gamma}$ , and which may be referred to as the invensor, the reversor, the conjugator, respectively. One can then utilize the expansion, valid for any operator  $b$ , and for each  $\Omega$ ,

$$(3.18) \quad e^{i\alpha\Omega} b e^{-i\alpha\Omega} = b + i\alpha[\Omega, b] + \frac{(i\alpha)^2}{2!}[\Omega, [\Omega, b]] + \frac{(i\alpha)^3}{3!}[\Omega, [\Omega, [\Omega, b]]] + \dots$$

For example, suppose one can find an operator  $\Omega'$  for which

$$(3.19) \quad [\Omega', b_{\mathbf{k}}] = -i\eta b_{-\mathbf{k}} \quad \text{and} \quad [\Omega', b_{-\mathbf{k}}] = i\eta^* b_{\mathbf{k}}; \quad \eta\eta^* = 1;$$

then

$$(3.20) \quad [\Omega', [\Omega', b_{\mathbf{k}}]] = -i\eta[\Omega', b_{-\mathbf{k}}] = \eta\eta^* b_{\mathbf{k}} = b_{\mathbf{k}}$$

$$(3.21) \quad [\Omega', [\Omega', [\Omega', b_{\mathbf{k}}]]] = [\Omega', b_{\mathbf{k}}] = -i\eta b_{-\mathbf{k}} \quad \text{etc.},$$

and one can write

$$(3.22) \quad e^{i\alpha\Omega'} b_{\mathbf{k}} e^{-i\alpha\Omega'} = b_{\mathbf{k}} \left\{ 1 + \frac{(i\alpha)^2}{2!} + \dots \right\} - i\eta b_{-\mathbf{k}} \left\{ i\alpha + \frac{(i\alpha)^3}{3!} + \dots \right\} \\ = b_{\mathbf{k}} \cos \alpha + \eta b_{-\mathbf{k}} \sin \alpha.$$

Putting now  $\alpha = \pi/2$  one obtains

$$(3.23) \quad e^{i(\pi/2)\Omega'} b_{\mathbf{k}} e^{-i(\pi/2)\Omega'} = \eta b_{-\mathbf{k}}$$

which is just the type of relation wanted by the first equation (3.11). It is clear that equations (3.19) cannot apply to the entire  $\mathbf{k}$ -space, because changing  $\mathbf{k}$  into  $-\mathbf{k}$  in the first equation does not give the second equation. However, if one considers a certain half space of  $\mathbf{k}$ , say the half space  $k_3 > 0$ , then one can satisfy equations (3.19) and therefore (3.23) with the obviously Hermitean operator

$$(3.24) \quad \Omega' = \sum'_{\mathbf{k}'} \left( i\eta b_{\mathbf{k}'}^+ b_{-\mathbf{k}'} - i\eta^* b_{-\mathbf{k}'}^+ b_{\mathbf{k}'} \right).$$

The prime on the summation sign means the sum is to be taken over the half space of  $\mathbf{k}'$  only, so that  $\delta_{-\mathbf{k}', \mathbf{k}} = \delta_{\mathbf{k}', -\mathbf{k}} = 0$  for all  $\mathbf{k}$  with  $k_3 > 0$ . The other half space of  $\mathbf{k}$  is not taken care of by this, because if one considers now all  $\mathbf{k}$  with  $k_3 < 0$ , so that  $\delta_{\mathbf{k}', \mathbf{k}} = \delta_{-\mathbf{k}', -\mathbf{k}} = 0$  for all vectors  $\mathbf{k}'$  appearing in the sum (3.24), one obtains

$$(3.25) \quad [\Omega', b_{\mathbf{k}}] = i\eta^* b_{-\mathbf{k}} \quad \text{and} \quad [\Omega', b_{-\mathbf{k}}] = -i\eta b_{\mathbf{k}} \quad (k_3 < 0)$$

which gives again (3.23) in the same half space  $-\mathbf{k}(k_3 < 0)$  or  $\mathbf{k}(k_3 > 0)$  as above. In order to obtain

$$(3.26) \quad e^{i(\pi/2)\Omega''} b_{-\mathbf{k}} e^{-i(\pi/2)\Omega''} = \eta b_{\mathbf{k}}$$

for  $k_3 > 0$ , one needs another operator  $\Omega''$  satisfying

$$(3.27) \quad [\Omega'', b_{-\mathbf{k}}] = -i\eta b_{\mathbf{k}} \quad \text{and} \quad [\Omega'', b_{\mathbf{k}}] = i\eta^* b_{-\mathbf{k}} \quad (k_3 > 0).$$

These relations are true for

$$(3.28) \quad \Omega'' = \sum'_{\mathbf{k}'} \left( i\eta b_{-\mathbf{k}'}^+ b_{\mathbf{k}'} - i\eta^* b_{\mathbf{k}'}^+ b_{-\mathbf{k}'} \right)$$

where the summation goes over the same half space  $\mathbf{k}'$  as in (3.24).

Expressions (3.23) and (3.26) can be consolidated into one formula, applicable to the entire  $\mathbf{k}$ -space, by writing  $\Pi$  as a direct product, so that

$$(3.29) \quad e^{i(\pi/2)\Omega} b_{\mathbf{k}} e^{-i(\pi/2)\Omega} = \eta b_{-\mathbf{k}}; \quad \Omega = \theta(\mathbf{k})\Omega' + \theta(-\mathbf{k})\Omega''$$

where  $\theta(\mathbf{k})$  is the singular operator

$$(3.30) \quad \theta(\mathbf{k}) = \begin{cases} 1 & \text{for } k_3 > 0 \\ 0 & \text{for } k_3 < 0 \end{cases} \quad \text{with } \theta(\mathbf{k}) + \theta(-\mathbf{k}) = 1 \text{ for all } \mathbf{k}.$$

Applying this kind of argument one finds the following operator  $\Pi$  satisfying the relations (3.11) and (3.4):

$$(3.31) \quad \Pi = e^{i(\pi/2)\Omega_{\Pi}}; \quad \Omega_{\Pi} = \theta(\mathbf{k})\Omega'_{\Pi} + \theta(-\mathbf{k})\Omega''_{\Pi}$$

where

$$(3.32) \quad \Omega'_{\Pi} = -\Omega''_{\Pi} = i\eta_{\Pi} \sum'_{\mathbf{k}'} (B_{\mathbf{k}'} - B_{-\mathbf{k}'})$$

with

$$(3.33) \quad B_{\mathbf{k}} = b_{\mathbf{k}}^+ (1) b_{-\mathbf{k}} (1) - b_{-\mathbf{k}}^+ (2) b_{\mathbf{k}} (2).$$

An analogous expression is obtained for the operator  $U$  satisfying the relations (3.14):

$$(3.34) \quad U = e^{i(\pi/2)\Omega_{\Theta}}; \quad \Omega_{\Theta} = \theta(\mathbf{k})\Omega'_{\Theta} + \theta(-\mathbf{k})\Omega''_{\Theta}$$

where

$$(3.35) \quad \Omega'_{\Theta} = \Omega''_{\Theta}^* = i \sum'_{\mathbf{k}'} (\eta_{\Theta} B_{\mathbf{k}'} - \eta_{\Theta}^* B_{-\mathbf{k}'}).$$

The transformation equations (3.13) for the particle conjugation operator require an operator  $\Omega'_{\Gamma}$  such that

$$(3.36) \quad [\Omega'_{\Gamma}, b_{\mathbf{k}} (1)] = -i\eta_{\Gamma}^* b_{\mathbf{k}} (2); \quad [\Omega'_{\Gamma}, b_{\mathbf{k}} (2)] = i\eta_{\Gamma} b_{\mathbf{k}} (1)$$

yielding by the argument employed in (3.22) and (3.23) the desired relation

$$(3.37) \quad e^{i(\pi/2)\Omega'_{\Gamma}} b_{\mathbf{k}} (1) e^{-i(\pi/2)\Omega'_{\Gamma}} = \eta_{\Gamma}^* b_{\mathbf{k}} (2),$$

and an operator  $\Omega''_\Gamma$  such that

$$(3.38) \quad [\Omega''_\Gamma, b_{\mathbf{k}}(2)] = -i\eta_\Gamma b_{\mathbf{k}}(1); \quad [\Omega''_\Gamma, b_{\mathbf{k}}(1)] = i\eta_\Gamma^* b_{\mathbf{k}}(2)$$

yielding the other desired relation

$$(3.39) \quad e^{i(\pi/2)\Omega''_\Gamma} b_{\mathbf{k}}(2) e^{-i(\pi/2)\Omega''_\Gamma} = \eta_\Gamma b_{\mathbf{k}}(1).$$

Obviously, equations (3.38) cannot be obtained from (3.36) with the same  $\Omega$  by interchange of the particle labels 1 and 2, this essential asymmetry in  $r$ -space is the analog of the asymmetry in  $\mathbf{k}$ -space encountered for the operator  $\Omega_\Pi$ . It is readily verified that the following operators satisfy the equations (3.36) and (3.38) respectively:

$$(3.40) \quad \Omega'_\Gamma = i \sum_{\mathbf{k}'} \{ \eta_{\mathbf{k}'}^* b_{\mathbf{k}'}^+(1) b_{\mathbf{k}'}^-(2) - \eta_\Gamma b_{\mathbf{k}'}^+(2) b_{\mathbf{k}'}^-(1) \}$$

$$(3.41) \quad \Omega''_\Gamma = -\Omega'_\Gamma$$

where the summation goes over the entire  $\mathbf{k}'$ -space. Introduction of the singular operator  $\theta$  allows one to consolidate the equations (3.37) and (3.39) into a single expression for the operator  $\Gamma$ ,

$$(3.42) \quad \Gamma = e^{i(\pi/2)\Omega_\Gamma}; \quad \Omega_\Gamma = \theta\{(-1)^{r+1}\}\Omega'_\Gamma + \theta\{(-1)^r\}\Omega''_\Gamma$$

where  $r$  is the particle label of the operator  $b_{\mathbf{k}}(r)$  to which  $\Gamma$  is applied.

As an example of how to manipulate the singular operator  $\theta$ , consider the expression

$$(3.43) \quad \Pi^2 = e^{i(\pi/2)\{\theta(\mathbf{k})\Omega'_\Pi + \theta(-\mathbf{k})\Omega''_\Pi\}} e^{i(\pi/2)\{\theta(\mathbf{k})\Omega'_\Pi + \theta(-\mathbf{k})\Omega''_\Pi\}}.$$

Since the operator  $\theta(\mathbf{k})$  in the second exponential selects the index  $\mathbf{k}$  ( $k_3 > 0$ ) in any expression to which it is applied, and since at the same time  $\Omega'_\Pi$  converts  $\mathbf{k}$  into  $-\mathbf{k}$ , this part of the second exponential will always necessarily be bunched together with the term  $\theta(-\mathbf{k})\Omega''_\Pi$  appearing in the first exponential, and similarly the term with  $\theta(-\mathbf{k})$  in the second exponential can only be bunched with the first term containing  $\theta(\mathbf{k})$  in the first exponential. The operation  $\Pi^2$  is therefore identical with

$$(3.44) \quad \Pi^2 = e^{i(\pi/2)\{(\Omega''_\Pi + \Omega'_\Pi)\theta(-\mathbf{k}) + (\Omega'_\Pi + \Omega''_\Pi)\theta(\mathbf{k})\}} = e^{i(\pi/2)(\Omega'_\Pi + \Omega''_\Pi)}.$$

Since according to (3.32)  $\Omega'_\Pi + \Omega''_\Pi = 0$  one has  $\Pi^2 = 1$  as expected. Similarly one finds

$$(3.45) \quad UU^* = e^{i(\pi/2)(\Omega'_\Theta - \Omega''_\Theta)}$$

which is the identity operation because of (3.35), and

$$(3.46) \quad \Gamma^2 = e^{i(\pi/2)(\Omega'_\Gamma + \Omega''_\Gamma)}$$

which is the identity operation because of (3.41).

#### 4. REPRESENTATION OF $\Pi$ , $\Theta$ , $\Gamma$ FOR A SPINOR FIELD

The operators of space inversion, time reversal, and particle conjugation for the spin- $\frac{1}{2}$  field (2.9) are defined by

$$(4.1) \quad \Pi\psi(\mathbf{x}, t)\Pi^{-1} = i\eta_\Pi \Lambda_\Pi \psi(-\mathbf{x}, t)$$

$$(4.2) \quad \Theta\psi(\mathbf{x},t)\Theta^{-1} = \eta_{\Theta}^* \Lambda_{\Theta} \psi^+(\mathbf{x},-t)$$

$$(4.3) \quad \Gamma\psi(\mathbf{x},t)\Gamma^{-1} = \eta_{\Gamma}^* \Lambda_{\Gamma} \tilde{\psi}(\mathbf{x},t)$$

where the matrices  $\Lambda$  have already been given in equations (2.14)–(2.16). In contrast to the case of the boson field one cannot impose here the condition that each of these operations applied twice should give the identity operation, because  $\psi$  itself is not an observable. Regarding  $\Pi$  there is an ambiguity in phase, first noticed by Yang and Tiomno (1950), which will be settled here by defining  $\Pi$  such that

$$(4.4) \quad \Pi^2 = -1.$$

This yields for the factor  $\eta_{\Pi}$  the condition

$$(4.5) \quad \eta_{\Pi}^2 = 1.$$

Regarding  $\Theta$  it follows from the definition (4.2), as will be shown below, that the four-component spinor field in conjunction with the assumption that  $\eta_{\Theta}$  be a complex, and not a hypercomplex, number requires

$$(4.6) \quad \Theta^2 = -1$$

which imposes on  $\eta_{\Theta}$  the relation

$$(4.7) \quad \eta_{\Theta} \eta_{\Theta}^* = 1.$$

Regarding  $\Gamma$  it is similarly found that if  $\eta_{\Gamma}$  should be a complex number, then

$$(4.8) \quad \Gamma^2 = +1$$

yielding for the phase  $\eta_{\Gamma}$  the relation

$$(4.9) \quad \eta_{\Gamma} \eta_{\Gamma}^* = 1.$$

The existence of other unusual types of spinor fields requiring different conditions on the squares of  $\Pi$ ,  $\Theta$ ,  $\Gamma$  poses a number of interesting questions, but these are outside the scope of the present work. The reader is referred to the paper of Shirokov (1960).

Writing down the equations (4.1)–(4.3) in terms of the expansions (2.9) and (2.12), applying the definitions (2.14)–(2.16), and changing the summation index on the right-hand sides of (4.1) and (4.2) from  $\mathbf{k}$  to  $-\mathbf{k}$  yields upon comparison of coefficients the following relations which give the effect of  $\Pi$ ,  $\Theta$ ,  $\Gamma$  on the annihilation and creation operators:

$$(4.10) \quad \Pi a_{\mathbf{k}}(s,1)\Pi^{-1} = i\eta_{\Pi} a_{-\mathbf{k}}(s,1); \quad \Pi a_{\mathbf{k}}(s,2)\Pi^{-1} = i\eta_{\Pi} a_{-\mathbf{k}}(s,2)$$

$$(4.11) \quad \Theta a_{\mathbf{k}}(s,1)\Theta^{-1} = (-1)^s \eta_{\Theta}^* a_{-\mathbf{k}}^+(s',1); \quad \Theta a_{\mathbf{k}}(s,2)\Theta^{-1} = (-1)^s \eta_{\Theta} a_{-\mathbf{k}}^+(s',2) \quad \left. \right\}$$

$$(4.12) \quad \Gamma a_{\mathbf{k}}(s,1)\Gamma^{-1} = (-1)^s \eta_{\Gamma}^* a_{\mathbf{k}}(s',2); \quad \Gamma a_{\mathbf{k}}(s,2)\Gamma^{-1} = (-1)^{s'} \eta_{\Gamma} a_{\mathbf{k}}(s',1) \quad \left. \right\} \quad s' \neq s.$$

In the derivation of (4.10) use has been made of the reality of  $\eta_{\Pi}$ ,  $\eta_{\Pi}^* = \eta_{\Pi}$ . To verify the statements made in connection with the squares of the operations  $\Theta$  and  $\Gamma$  apply  $\Theta$  once more to (4.11) and  $\Gamma$  once more to (4.12). One finds

immediately  $\Theta^2 a_{\mathbf{k}}(s, r) \Theta^{-2} = -\eta_{\Theta}^* \eta_{\Theta} a_{\mathbf{k}}(s, r)$  and  $\Gamma^2 a_{\mathbf{k}}(s, r) \Gamma^{-2} = +\eta_{\Gamma}^* \eta_{\Gamma} a_{\mathbf{k}}(s, r)$ . It is therefore not possible to satisfy  $\Theta^2 = +1$ ,  $\Gamma^2 = -1$  with complex numbers  $\eta_{\Theta}$  and  $\eta_{\Gamma}$ .

As in the case of the boson field the antiunitary operator  $\Theta$  is conveniently decomposed into a unitary operator  $U$  and the nonlinear operator of transposition  $K$ , so that

$$(4.13) \quad \begin{aligned} U a_{\mathbf{k}}(s, 1) U^{-1} &= (-1)^s \eta_{\Theta} a_{-\mathbf{k}}(s', 1), \\ U a_{\mathbf{k}}(s, 2) U^{-1} &= (-1)^s \eta_{\Theta}^* a_{-\mathbf{k}}(s', 2); \end{aligned} \quad s' \neq s.$$

In the derivation of (4.13) from (4.11) with the definition  $\Theta = UK$  use has been made of the reality of the operators  $a$ , which may be represented by matrices satisfying  $a^+ = \tilde{a}$ .

For the actual construction of  $\Pi$ , consider an operator defined in the half space  $k_3 > 0$

$$(4.14) \quad \Pi'_{\mathbf{k}} = 1 - a_{\mathbf{k}}^+ a_{\mathbf{k}} - a_{-\mathbf{k}}^+ a_{-\mathbf{k}} + i\eta(a_{\mathbf{k}}^+ a_{-\mathbf{k}} - a_{-\mathbf{k}}^+ a_{\mathbf{k}}) \quad (k_3 > 0).$$

By application of the anti C.R.'s (2.20) one finds the relations

$$(4.15) \quad \Pi'_{\mathbf{k}} a_{\mathbf{k}} = i\eta a_{-\mathbf{k}} \Pi'_{\mathbf{k}}; \quad \Pi'_{\mathbf{k}} a_{\mathbf{k}}^+ = -i\eta a_{-\mathbf{k}}^+ \Pi'_{\mathbf{k}} \quad (k_3 > 0)$$

which are of the form desired by equation (4.10). If one applies (4.14) to  $a_{-\mathbf{k}}$  ( $k_3 > 0$ ), however, one obtains relations which differ in sign from (4.15). It is therefore necessary to introduce another operator

$$(4.16) \quad \Pi''_{\mathbf{k}} = 1 - a_{\mathbf{k}}^+ a_{\mathbf{k}} - a_{-\mathbf{k}}^+ a_{-\mathbf{k}} - i\eta(a_{\mathbf{k}}^+ a_{-\mathbf{k}} - a_{-\mathbf{k}}^+ a_{\mathbf{k}}) \quad (k_3 > 0)$$

which yields the desired relations

$$(4.17) \quad \Pi''_{\mathbf{k}} a_{-\mathbf{k}} = i\eta a_{\mathbf{k}} \Pi''_{\mathbf{k}}; \quad \Pi''_{\mathbf{k}} a_{-\mathbf{k}}^+ = -i\eta a_{\mathbf{k}}^+ \Pi''_{\mathbf{k}} \quad (k_3 > 0).$$

Introduction of the singular operator  $\theta(\mathbf{k})$  as in (3.30) allows one to consolidate (4.15) and (4.17) into equation (4.10), applicable to the entire  $\mathbf{k}$ -space, by writing

$$(4.18) \quad \Pi = \Pi' \theta(\mathbf{k}) + \Pi'' \theta(-\mathbf{k})$$

where

$$(4.19) \quad \Pi' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 \prod_{r=1}^2 \Pi'_{\mathbf{k}'}(s, r); \quad \Pi'' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 \prod_{r=1}^2 \Pi''_{\mathbf{k}'}(s, r)$$

and

$$(4.20) \quad \Pi'_{\mathbf{k}}(s, r) = \Pi''_{\mathbf{k}}(s, r) = 1 - N_{\mathbf{k}}(s, r) - N_{-\mathbf{k}}(s, r) + i\eta_{\Pi} \{A_{\mathbf{k}}(s, r) - A_{-\mathbf{k}}(s, r)\}$$

with

$$(4.21) \quad N_{\mathbf{k}}(s, r) = a_{\mathbf{k}}^+(s, r) a_{\mathbf{k}}(s, r); \quad A_{\mathbf{k}}(s, r) = a_{\mathbf{k}}^+(s, r) a_{-\mathbf{k}}(s, r).$$

The prime on the product sign  $\prod'$  indicates that the factors are to be taken from the half space  $\mathbf{k}'$  with  $k'_3 > 0$  only.

The construction of  $U$  proceeds along the same line of argument, it is only slightly more complicated by the fact that relations (4.13) distinguish between particle labels  $r = 1$  and  $r = 2$ , because in general  $\eta_0 \neq \eta_0^*$ . One finds

$$(4.22) \quad U = U' \theta(\mathbf{k}) + U'' \theta(-\mathbf{k})$$

where

$$(4.23) \quad U' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 U'_{\mathbf{k}'}(s,1) U'_{\mathbf{k}'}(s,2); \quad U'' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 U''_{\mathbf{k}'}(s,1) U''_{\mathbf{k}'}(s,2)$$

and

$$(4.24a) \quad U'_{\mathbf{k}}(s,1) = U''_{-\mathbf{k}}(s,1) = 1 - N_{\mathbf{k}}(s,1) - N_{-\mathbf{k}}(s',1) \\ + (-1)^s \{ \eta_0 C_{\mathbf{k}}(s,1) + \eta_0^* C_{-\mathbf{k}}(s',1) \} \quad \left. \right\} \quad s' \neq s$$

$$(4.24b) \quad U'_{\mathbf{k}}(s,2) = U''_{-\mathbf{k}}(s,2) = 1 - N_{\mathbf{k}}(s,2) - N_{-\mathbf{k}}(s',2) \\ + (-1)^s \{ \eta_0^* C_{\mathbf{k}}(s,2) + \eta_0 C_{-\mathbf{k}}(s',2) \} \quad \left. \right\} \quad s' \neq s$$

with

$$(4.25) \quad C_{\mathbf{k}}(s,r) = a_{\mathbf{k}}^+(s,r) a_{-\mathbf{k}}(s',r); \quad s' \neq s.$$

Finally, one finds for the operator  $\Gamma$  satisfying the relations (4.12) the representation

$$(4.26) \quad \Gamma = \Gamma' \theta\{(-1)^{r+1}\} + \Gamma'' \theta\{(-1)^r\}$$

where

$$(4.27) \quad \Gamma' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 \Gamma_{\mathbf{k}'}(s,1); \quad \Gamma'' = \prod'_{\mathbf{k}'} \prod_{s=1}^2 \Gamma_{\mathbf{k}'}(s,2)$$

and

$$(4.28a) \quad \Gamma_{\mathbf{k}}(s,1) = 1 - N_{\mathbf{k}}(s,1) - N_{\mathbf{k}}(s',2) \\ + (-1)^s \{ \eta_{\Gamma} D_{\mathbf{k}}(s,1) + \eta_{\Gamma}^* D_{\mathbf{k}}(s',2) \} \quad \left. \right\} \quad s' \neq s$$

$$(4.28b) \quad \Gamma_{\mathbf{k}}(s,2) = 1 - N_{\mathbf{k}}(s,2) - N_{\mathbf{k}}(s',1) \\ + (-1)^{s'} \{ \eta_{\Gamma} D_{\mathbf{k}}(s,2) + \eta_{\Gamma}^* D_{\mathbf{k}}(s',1) \} \quad \left. \right\} \quad s' \neq s$$

with

$$(4.29) \quad D_{\mathbf{k}}(s,r) = a_{\mathbf{k}}^+(s,r) a_{\mathbf{k}}(s',r'); \quad s' \neq s; \quad r' \neq r.$$

In (4.27) the product  $\prod'$  goes now over the entire  $\mathbf{k}'$ -space, and the asymmetry is restricted, as in the corresponding boson field case, to the particle label  $r$  of any operator or state on which  $\Gamma$  operates.

It seems worth noting that the operators  $\Pi$ ,  $U$ ,  $\Gamma$  for spinor fields may be written as exponentials, in analogy to the corresponding cases for tensor fields. For example, considering again the operator (4.14)

$$(4.30) \quad \Pi'_{\mathbf{k}} = 1 + \Omega'_{\mathbf{k}}$$

one finds easily that  $\Omega'_{\mathbf{k}}$ , by virtue of the anti C.R.'s (2.20) has the property

$$(4.31) \quad (\Omega'_{\mathbf{k}})^n = (-2)^{n-1} \Omega'_{\mathbf{k}}.$$

This allows one to write

$$(4.32) \quad \Pi'_{\mathbf{k}} = e^{i(\pi/2)\Omega'_{\mathbf{k}}} = 1 + \frac{\Omega'_{\mathbf{k}}}{2} \{i \sin \pi + (1 - \cos \pi)\} = 1 + \Omega'_{\mathbf{k}}.$$

Similar relations are found for the other operators  $U_{\mathbf{k}}$  and  $\Gamma_{\mathbf{k}}$ . In practical applications it appears, however, more convenient to use the already expanded bilinear forms (4.20), (4.24), and (4.28).

#### ACKNOWLEDGMENTS

Mr. K. H. Mariwalla has checked some of the calculations and pointed out an error in an early draft of this paper. Prof. W. Opechowski kindly drew the author's attention to the work of Watanabe.

#### REFERENCES

BOGOLIUBOV, N. N. and SHIRKOV, D. V. 1959. Introduction to the theory of quantized fields (Interscience Publishers, New York).

CORINALDESI, E. 1958. Nuclear Phys. **7**, 305.

HAMILTON, J. 1959. The theory of elementary particles (Oxford University Press).

JAUCH, J. M. and ROHRLICH, F. 1955. The theory of photons and electrons (Cambridge, Mass.).

KÄLLEN, G. 1958. Encyclopedia of physics, Vol. V(1) (Springer-Verlag, Berlin).

LÜDERS, G. 1954. Danske Videnskab. Selskab, **28**, No. 5.

— 1957. Ann. Phys. **2**, 1.

MANDL, F. 1959. Introduction to quantum field theory (Interscience Publishers, New York).

NIGAM, B. P. and FOLDY, L. L. 1956. Phys. Rev. **102**, 1410.

SCHWINGER, J. 1951. Phys. Rev. **82**, 914.

SHIROKOV, YU. M. 1960. Nuclear Phys. **15**, 1.

WATANABE, S. 1955. Revs. Modern Phys. **27**, 40.

WICK, G. 1958. Ann. Rev. Nuclear Sci. **8**, 1.

WOLFENSTEIN, L. and RAVENHALL, D. G. 1952. Phys. Rev. **88**, 279.

YANG, C. N. and TIOMNO, J. 1950. Phys. Rev. **79**, 495.

# HIGH-FREQUENCY RADIO-WAVE BLACK-OUTS AT MEDIUM AND HIGH LATITUDES DURING A SOLAR CYCLE<sup>1</sup>

C. COLLINS, D. H. JELLY, AND A. G. MATTHEWS

## ABSTRACT

Recent studies at the Defence Research Telecommunications Establishment, Ottawa, comparing v.h.f. riometer and h.f. ionosonde data for the International Geophysical Year, have shown that the two types of absorption events, 'polar cap' and 'auroral', can be identified in the occurrence patterns of ionosonde black-outs. A study based on this comparison has been made of all black-outs observed at a number of medium- and high-latitude ionosonde stations during the period 1949 to 1959. It has been found that the two kinds of absorption events show markedly different variations during the 11 years. Several other temporal and spatial features of the phenomena are also discussed.

## INTRODUCTION

The complete absorption of high-frequency radio waves in the disturbed ionosphere is a well-known phenomenon. The effect, commonly termed 'black-out', has been the subject of both detailed and synoptic investigations in several countries over the past twenty years. One of the first advances in this field was made by Wells (1942), who showed the close association between short-duration black-outs and magnetic bays at College, Alaska. Later statistical studies (Lindquist 1951; Agy 1954a, b; Cox and Davies 1954) indicated some of the large-scale temporal and spatial characteristics of the black-out events. Meek (1952), in his studies of arctic ionospheric disturbances, also described some of the gross features of the absorption phenomena and showed the close association between black-outs and other manifestations of the high-latitude disturbances. A summary of the early work in this field has been given in the comprehensive review published by Little, Rayton, and Roof (1956).

Much of this work was based on the black-out reports from vertical-incidence ionospheric sounding stations. The results led several workers to suggest that there might be more than one kind of black-out condition although no classifications were adopted. This was perhaps due to the nature of the ionosonde black-out data which gave only a qualitative picture of the absorbing regions. More recently, quantitative investigations of the absorption phenomena have been made by a number of groups using measurements of cosmic noise at very high frequencies (20-60 Mc/s). The results of these investigations have confirmed the belief of the earlier workers by showing that the high-latitude absorption events are mainly of two kinds. These are now generally termed 'polar-cap absorption' and 'auroral absorption'.

Several of the polar-cap absorption events have been reported in detail (Bailey 1957, 1959; Hakura, Takenoshita, and Otsuki 1958; Reid and Collins

<sup>1</sup>Manuscript received August 15, 1960.

Contribution from the Radio Physics Laboratory, Defence Research Telecommunications Establishment, Defence Research Board, Ottawa, Canada. Work carried out under Project No. D48-95-11-36.

1959; and others). In their paper, Reid and Collins divided the abnormal absorption events into three classes depending on the characteristics of the cosmic noise records and on the association with magnetic and auroral activity. The classes were designated as types 1, 2, and 3. Type 1 referred to the well-known sudden ionospheric disturbance (SID). The authors did not consider this type of event relevant to their study, and, since Cox and Davies (1954) have shown that SID's during the 4-year period 1949-52 did not contribute significantly to medium- and high-latitude black-outs, the phenomenon will not be discussed in this paper. Type 2 produced a characteristically irregular trace on the riometer records and the events were easily identified. They occurred at both Ottawa and Churchill, although not always simultaneously, and they often coincided with the occurrence of visible aurora and local magnetic activity. It is probable that these events were similar to the short-duration black-outs studied by Wells. This type 2 is now known as auroral absorption. Although there is considerable evidence to show that there is a type of absorption occurring at geomagnetic latitudes where visible aurora is usually observed (about 55 to 75 degrees) the precise relationship between the radio and the optical phenomena is not yet understood and the term 'auroral' must be used with some caution. In the present paper it is used in conjunction with 'absorption' and 'black-out' to designate occurrences closely associated in time with magnetic disturbances in or near the zone of visible aurora.

Type 3 absorption has been more clearly defined. It began a few hours after the start of a large solar flare and produced a much smoother variation on the riometer records than did the auroral type. It persisted for a number of days and showed a much greater intensity during the day than during the night. No magnetic disturbance was associated with the early hours of an event although a large magnetic storm usually occurred 10 to 30 hours after the start of the absorption. The absorbing region, it was subsequently learned, extended over most of the polar cap (Hakura *et al.* 1958) and hence the name 'polar-cap absorption' came to be associated with it. It is considered to be mainly a high-latitude phenomenon although, as results from the present study show, the effects are frequently observed at geomagnetic latitudes as far south as Winnipeg (59.6°).

Both these types of absorption are detected by the h.f. ionosondes and can be identified if the cosmic noise measurements are used to interpret the vertical-incidence ionosonde measurements. There are two obvious advantages in this procedure. First, since many more years of ionosonde data than cosmic noise data are available, the combination of the two makes possible the determination of the long-term variations of the events observed on the cosmic noise records. Second, the h.f. black-out, which is at best a crude measure of abnormal absorption, now possesses more physical significance for the statistical and synoptic studies of ionospheric disturbances. The procedure is only justified, however, if a close correlation can be established between the results obtained by the two kinds of observations. This point is considered in detail in the following section of the paper.

In the present study of black-out occurrences both ionosonde and cosmic

noise measurements have been used and the foregoing classifications from the v.h.f. studies have been applied to the analysis of the h.f. black-out data. All black-out occurrences at a number of Canadian ionosondes for the period 1949 to 1959 have been considered. This 11-year period was sufficiently long to reveal significant differences in the large-scale temporal features of the two kinds of black-out, suggesting that they might be associated with different types of disturbances on the sun.

#### ANALYSIS OF DATA

Data for the study were collected at the stations listed in Table I. The h.f. ionosondes were conventional instruments and their characteristics have been

TABLE I  
Co-ordinates and time zones of stations

Station	Geographic		Geomagnetic	
	Latitude north $\phi^\circ$	Longitude west $\lambda^\circ$	North $\phi^\circ$	L.M.T. $^{\circ}\text{W}$
Alert	82.6	62.6	85.0	75
Eureka	80.0	85.9	86.5	75
Resolute Bay	74.7	94.9	82.9	90
Clyde River	70.5	68.6	81.9	75
Baker Lake	64.3	96.0	73.7	90
Churchill	58.8	94.2	68.8	90
Meenook	54.6	113.3	62.0	105
Winnipeg	49.9	97.4	59.6	90
Ottawa	45.4	75.7	56.9	75
Agincourt	43.8	79.3	55.0	75

described in detail in the Canadian I.G.Y. program (Meek 1957). The cosmic noise recorders were situated at Ottawa and Churchill, and operated on frequencies close to 30 Mc/s. Techniques of equipment operation and methods of data reduction have been described by Little and Leinbach (1958) and Reid and Collins (1959). The analysis of the data will be described in two parts, one dealing with polar-cap black-outs and the other with auroral black-outs.

#### (a) *Polar-cap Black-out*

The first step in the analysis was a comparison of polar-cap absorption events identified from the riometer measurements and those selected independently from ionosonde black-out occurrences. The I.G.Y. period was selected because data from all the stations were available for most of the period. Auroral and polar-cap absorption events were identified directly from the riometer records. It was found that the Churchill riometer in the 18-month period had detected 15 disturbances involving polar-cap absorption greater than 2 db. Some of these included short periods of auroral absorption in the later parts of the disturbances. No definite occurrence of polar-cap absorption was found on the Ottawa records, although auroral absorption was observed at that station quite frequently.

An independent examination was then made of the ionospheric data from eight stations—Resolute Bay, Baker Lake, and Ottawa for the full I.G.Y. period, and Eureka, Alert, Clyde River, Churchill, and Meanook for 1958. At these stations sweep-frequency soundings were made at 15-minute intervals and recorded in the usual way on  $f$ -plots. All black-outs of more than 3-hours' duration were plotted according to the geomagnetic latitude of the station and as a function of universal time (U.T.). It was then found that almost all the black-out occurrences were arranged in a number of distinctive periods of about 2 to 5 days. These black-out groupings showed marked similarities and were clearly separated by relatively long intervals containing no black-out. The starting and finishing times of these absorption events could be determined easily in most cases and 14 of these discrete events were found. A comparison then showed that the 14 long-enduring black-out periods selected from the ionosonde data corresponded to 14 of the 15 polar-cap absorption events identified from the Churchill riometer.

A further examination was made of the ionosonde data for the period of 22 to 25 September 1958, which had not exhibited the distinctive pattern in the black-out plots although the cosmic noise absorption had reached a value of 4 db. It was found that, during this period, the minimum frequencies reflected from the ionosphere at the northern stations were 5 to 6 Mc/s higher than the monthly median. This indicated a large increase in absorption. But the critical frequencies during the day were also found to be unusually high, 8 to 10 Mc/s. This would effectively extend the range of the ionosonde and the absorption was not sufficiently intense to cause a black-out at the higher frequencies.

An examination was also made of the magnetic activity during the 14 black-out events, and histograms of the planetary indices,  $K_p$ , were drawn on the latitude-time plots. The disturbance which occurred on 22–25 August 1958 is shown in Fig. 1 and is typical of the polar-cap events which showed both polar-cap and auroral absorption. For ease of comparison half-hourly cosmic noise absorption values ( $c$ ) in decibels for Churchill are also shown. Black-out occurrences are plotted for stations listed in Table I. Planetary  $K_p$  indices at the top of the figure indicate the degree of magnetic activity. The occurrence of the solar flare that was probably associated with the event is shown at the approximate time as is that of the magnetic sudden commencement (S.C.). The S.C. appears to follow the increase in the  $K_p$  index but this is due only to the coarseness of the  $K_p$ . The flare lasted from 1417 to 1717 U.T. and was described by the Dominion Observatory, Ottawa, as being of importance 3 and having heliographic co-ordinates  $21^\circ$  N. and  $8^\circ$  W. The first evidence of a black-out is seen to follow the start of the flare by about 3 hours and to precede the onset of the magnetic disturbance by about 32 hours. The black-out is almost continuous at Alert, Resolute Bay, and Clyde River, a fact that is attributed to the ionosphere always being sunlit at northern latitudes in August. Baker Lake and Churchill show daytime black-out and nighttime recovery. The black-out at Ottawa, which begins with the increase in magnetic activity and occurs irregularly at night, is more likely to be due to auroral absorption than to the polar-cap type.

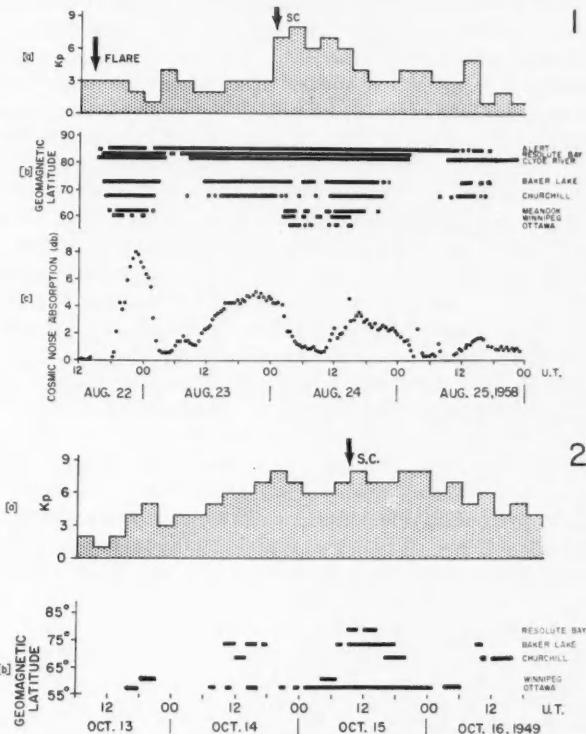


FIG. 1. Disturbance on 22-25 August, 1959, involving polar-cap and auroral absorption; (a) 3-hour planetary magnetic index  $K_p$ , (b) 15-minute black-outs observed at eight ionosonde stations, (c) variation in cosmic noise absorption at Churchill.

FIG. 2. Disturbance on 13-16 October, 1949, involving only auroral absorption; (a) 3-hour planetary magnetic index  $K_p$ , (b) hourly black-outs observed at five ionosonde stations.

The black-outs at Winnipeg and Meanook occurring on 22 August coincide with the beginning of polar-cap absorption and are probably of that type. Those occurring on 24 August, accompanied by an increase in  $K_p$ , are probably of the auroral type. The smoothness of the Churchill cosmic noise absorption indicates polar-cap type with the few irregularities (e.g. at 1430 on 24 August and 0300 on 25 August) being due to auroral type. It should be noted that there are no means of identifying the two types from the ionosonde data alone, but from the known characteristics of the phenomena one would expect only polar-cap absorption before the start of the magnetic disturbance. The simultaneous occurrence of the two types of absorption made it necessary to adopt a new term for the 14 long-enduring black-out periods identified in the ionosonde data. These will be referred to as 'polar-cap black-outs' in the remainder of this paper.

Most of the 14 polar-cap black-out events presented the same general configuration of black-out occurrences as that shown in Fig. 1. It was felt that

these patterns possessed sufficiently distinctive characteristics to permit their identification in earlier ionosonde data and the study was extended to the rest of the 11-year period (1949-1959).

Prior to the I.G.Y. most ionosondes took measurements for only a few minutes on the hour. The occurrence of black-out at the time of these hourly measurements was recorded in the  $f_{\text{min}}$  data sheets by the symbol  $B$ . This provides only a short sample of ionospheric conditions during an hour but, because of the large number of observations taken in the 11-year period, such measurements are adequate for a synoptic study of this kind. All  $B$ 's were plotted as in Fig. 1 from the monthly  $f_{\text{min}}$  tabulations for Resolute Bay, Baker Lake, Churchill, Winnipeg, and Ottawa. Several criteria were adopted for identifying the polar-cap black-out events. The black-out had to occur at more than one of the most northerly stations. It had to last more than 3 hours and be reasonably continuous for the duration of the disturbance. The lower latitude stations where the black-out might be discontinuous had to show some day-night variation. The pattern of occurrences for all the stations considered together had to show the same general configuration as presented in Fig. 1. Black-outs which did not satisfy these conditions were considered to be spurious or due to auroral absorption. In a few cases the variations of  $K_p$  were also used to confirm the selections. These criteria proved to be quite adequate and some 40 polar-cap black-outs were found to have occurred during the solar cycle. These events have been listed in Table II. In almost all cases the events were easily identified. There is a slight uncertainty about the exact number because a few of the long-enduring disturbances may have consisted of two or more overlapping events. These questionable cases are indicated in Table II.

#### *(b) Auroral Black-out*

After the periods of polar-cap black-out had been eliminated from the data a selection was made of all the  $B$ 's in the rest of the 11-year period which were believed to be due to auroral absorption. In order to identify these auroral  $B$ 's some criterion had to be adopted for their selection. This was necessary because the symbol  $B$  denotes only the absence of an echo from the ionosphere. The failure to detect an echo is usually due to the presence of an absorbing region but it may on occasions also be due to interference or very low critical frequencies of the reflecting region. An examination of the critical frequencies at some of the stations showed that many of the  $B$ 's did not signify real increases in absorption, particularly in years of low sunspot number. It was not considered practicable, however, to examine all the original records near the times of black-out. Such a procedure would be very laborious and would introduce a large subjective element into the analysis. It was therefore decided that the auroral  $B$ 's should be selected only if they occurred during a magnetically disturbed period.

Reference has already been made to Wells' study relating black-outs and magnetic bays. From observations at Saskatoon, Meek (1953) has published other detailed accounts of this association of h.f. absorption and magnetic

TABLE II  
Polar-cap black-outs identified

Year	Duration	Approx. starting time (U.T.)	Comments
1949	Jan. 23-26	1200	Possibly 2 events
	Apr. 11-14	1400	
	May 10-13	2000	
	June 4-7	1800	
	Aug. 3-8	1200	
	Nov. 19-21	1200	
1950	Feb. 2-3	0200	Possibly 2 events
	Feb. 22-24	1200	
	Mar. 27-29	1000	
	May 27-31	a.m.	
1951	Mar. 7-15	1600	Possibly 2 events
	Apr. 2-9	1600	
	June 11-19		
	Oct. 27-29	0800	
1952	No well-defined events.		
1953	Mar. 2-4	1800	Weak, indefinite
1954	No events		
1955	No events		
1956	Feb. 23-25	0600	
	Mar. 11-14	1600	
	Aug. 31-Sept. 2	1800	
	Nov. 13-15	1400	
1957	Jan. 21-23	1500	Weak Possibly 2 events (second starting on 22) Possibly 2 events (second starting on 3) Possibly 2 or 3 events
	Apr. 4-6	1200	
	June 20-26	1800	
	July 1-5	1200	
	Aug. 27-Sept. 4	1400	
	Sept. 21-23	1200	
	Oct. 21-22	1400	
1958	Feb. 10-11	p.m.	
	Mar. 14-15	2200	
	Mar. 25-28	1200-1600	
	Apr. 10-11	1400	
	July 7-10	0600	
	Aug. 16-18	1200	
	Aug. 22-25	1600	
	Aug. 26-27	a.m.	
1959	May 11-17	a.m.	
	July 10	1000	
	July 14-17		
	July 17-20		

NOTE: Starting times are estimated from beginning of black-out at several stations and are not an accurate measure of the starting time of each event.  
Possible weak events: 1955, Jan. 17-18; 1956, Apr. 27-28.

disturbance. Davies and Hagg (1955) have reported a pronounced correlation between nighttime ionospheric absorption on 2.0 Mc/s at Prince Rupert, and the local 3-hour range magnetic *K* index for values of *K* greater than 4. Heppner, Byrne, and Belon (1952) have reported that some parts of the

visible aurora at College were often accompanied by black-outs. It is almost certain that there would also be some magnetic activity at the time of such aurora. It therefore seemed reasonable in the present study to select the auroral black-outs by some level of magnetic disturbance. Since it cannot be shown that auroral black-outs are invariably accompanied by an increase in magnetic activity this method of selecting the data might reduce the total number of listed auroral black-out occurrences. This effect should be negligible, however, and most spurious black-out occurrences which were due only to the absence of a reflecting region should be excluded by this process. The criterion adopted was that the  $K$  index given by the nearest magnetic observatory should be greater than 3 in the 3-hour interval in which the black-out occurred. Indices from Agincourt, Meanook, and Resolute Bay were used although the data from Resolute Bay were available only for the years 1952-55.  $K$  was chosen for the purpose rather than the planetary index  $K_p$  as the former was considered to be more sensitive to local ionospheric disturbance.  $K_p$ , which indicates world-wide disturbance, has been plotted in Fig. 1 because black-outs from a number of stations have been shown.

Although many of the black-outs attributed to auroral absorption were randomly distributed, some did form distinct patterns. An example is shown in Fig. 2 for the disturbed period 13 to 16 October 1949. The magnetic activity is high. There are, however, no long-enduring black-outs at the high-latitude stations. The black-out occurrences are more sporadic and of shorter duration than during polar-cap black-outs and they are more frequent at Churchill, Winnipeg, and Ottawa. It can be seen that the general configuration of black-out occurrences is quite different from that shown in Fig. 1.

#### RESULTS OF SYNOPTIC SURVEY

The temporal variations of the absorption phenomena were examined in two ways. For the polar-cap black-out events both the number of events and the number of  $B$ 's making up the events were considered. In the case of auroral black-outs the periods designated as polar-cap black-outs were excluded because it was impossible to identify with certainty the type of absorption responsible for the black-out after the magnetic storm had started. In some of the figures all the occurrences of  $B$ 's have been shown except those making up the polar-cap black-out events, that is, all auroral  $B$ 's and all the spurious  $B$ 's rejected by the criterion of magnetic activity. These have been shown for several reasons which should be clear from the following descriptions of the figures.

##### (a) Secular Variation

The distribution by years of the 40 polar-cap black-outs is given by the histogram in Fig. 3. Shown on the same plot is the variation of the annual mean sunspot numbers published by the Central Radio Propagation Laboratory, Boulder, U.S.A. The close agreement of the two plots indicates that the occurrence of disturbances with polar-cap absorption is closely associated with the mean level of solar activity. Although the number of occurrences is too

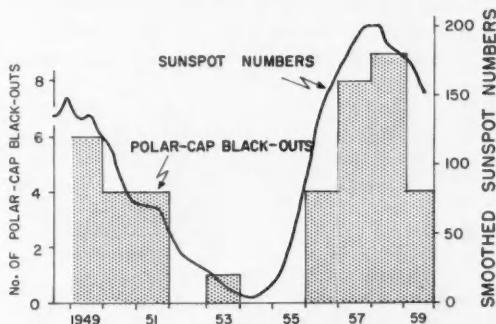


FIG. 3. Distribution of polar-cap black-outs during the sunspot cycle 1949-59.

few for a satisfactory statistical analysis, there is some indication that the outline of the histogram lags by about a year with respect to the sunspot number curve.

Yearly totals were also compiled of the numbers of *B*'s comprising the 40 polar-cap black-outs at Resolute Bay, Baker Lake, Churchill, and Winnipeg. These have not been reproduced here since the shape of the histograms was similar to that shown in Fig. 3. Some indication of the latitude dependence was, however, apparent in these histograms, with Resolute Bay having the greatest number of hourly black-outs and Winnipeg the least.

The secular variation of the auroral black-outs at five of the stations is shown in Fig. 4. For ease of comparison the sunspot number curve has been drawn on the Churchill and Winnipeg histograms. The dotted areas show the yearly totals of auroral *B*'s. The plain areas denote the number of *B*'s rejected by the criterion of magnetic activity. The years for which *K* indices were not available from Resolute Bay have been shaded with diagonal lines.  $\Phi$  denotes the geomagnetic latitude of the station. The vertical columns denoting the occurrences at Baker Lake and Winnipeg for 1959 have been drawn narrower than the others. In March 1959 the Baker Lake ionosonde was closed down and in September 1959 the transmitter power at Winnipeg was reduced from 10 kw to 1 kw. This loss of data was unfortunate but it does not seriously affect the results. It will readily be seen that there is some agreement between the outline of the histograms and the curve of solar activity. The variation in the black-out occurrences shows a definite lag on the variation in the sunspot number. Since these histograms are drawn to a common vertical scale they also show the relative frequency of occurrence of the auroral black-outs at the various stations. It can be seen that Baker Lake reported the fewest occurrences while Winnipeg observed the greatest number.

#### (b) Seasonal Variation

The histogram in Fig. 5 shows the seasonal variation of the polar-cap black-out events. Although the limited number of events does not allow much weight

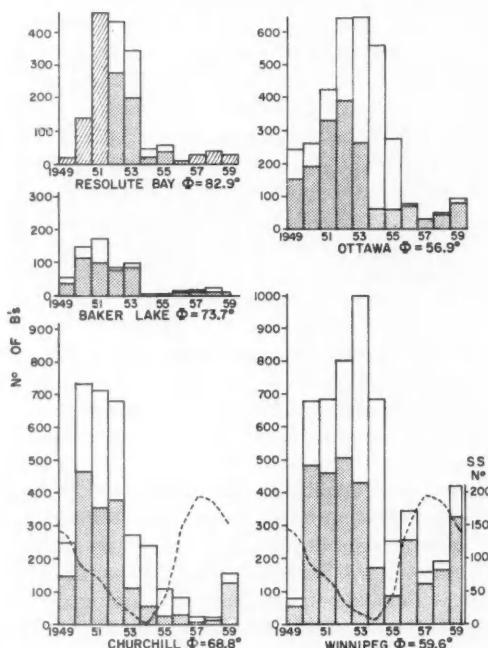


FIG. 4. Distribution of yearly totals of  $B$ 's at five stations during the solar cycle 1949-59, excluding periods of polar-cap black-out. Solid line histogram, all black-outs. Dotted areas, auroral black-outs. Diagonal line areas, periods for which  $K$  index was not available. Broken line, monthly mean sunspot numbers.  $\Phi$ , geomagnetic latitude.

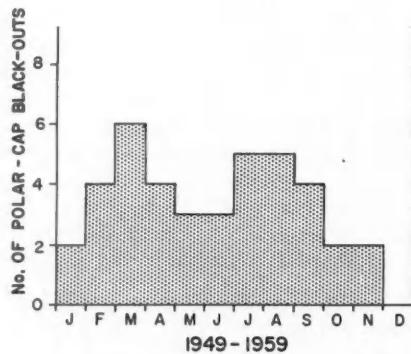


FIG. 5. Distribution of monthly totals of polar-cap black-outs for the years 1949-59.

to be attached to the result, the distribution does suggest slight increases in the occurrence somewhat before the equinoxes. It is perhaps interesting to note that throughout the 11-year period there were no events of this kind in December.

The seasonal variation for the number of *B*'s at Churchill and Ottawa is shown in Fig. 6. Part (a) gives the monthly total of the *B*'s comprising the

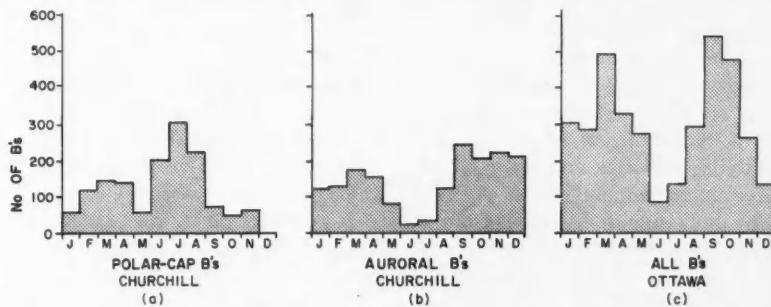


FIG. 6. Distributions of monthly totals of *B*'s; (a) polar-cap black-outs at Churchill, (b) auroral black-outs at Churchill, (c) all *B*'s at Ottawa, periods of polar-cap black-out included.

polar-cap black-outs at Churchill. This shows a maximum during the summer months. In contrast to this the distribution of auroral *B*'s for that station (b) shows a pronounced minimum for the summer months. It might be suggested that the summer maximum in the first histogram might be due to the occurrence of auroral black-outs which could not be extracted from the data. This seems unlikely, however, since the distribution of all black-outs for Ottawa also has a minimum about June and July as shown in (c) and no data were rejected for that station. It is well known that the polar-cap absorption is more intense when the lower ionosphere is sunlit and the maximum in the polar-cap distribution is probably due to the longer periods of sunlight at the higher latitudes during the summer. Plots were also made of the occurrences at the other stations. The histograms for Baker Lake and Winnipeg are similar to those shown in Fig. 6(a) and (b). For Resolute Bay, where the black-outs are mainly of the polar-cap type the histogram in part (a) predominates and the remaining black-outs show no significant seasonal variation. For Ottawa, which is south of the polar-cap absorption region for all but the most intense disturbances, one obtains a distribution of auroral *B*'s which is very similar to but slightly smaller than the histogram given in part (c).

#### (c) Diurnal Variation

Diurnal variations for the auroral black-outs at each of the five stations were derived by totalling the number of *B*'s for each hour during September and October for the 11 years. These have not been reproduced here since they differ only slightly from those already published by Cox and Davies (1954).

The hours (local mean time) of maximum occurrence at Ottawa, Winnipeg, Churchill, and Baker Lake were 0300, 0600, 0700, and 0800 respectively. No significant peak was found in the plot of the occurrences at Resolute Bay.

(d) *27-day Recurrence Tendencies*

In order to examine the data for 27-day recurrence patterns, an index was compiled using three categories according to the number of *B*'s during the day: (1) no *B*'s, (2) 3 hours or less, and (3) greater than 3 hours. Polar-cap black-outs from Churchill were plotted on one graph, all other *B*'s from the same station on another plot. The inclusion of all *B*'s both auroral and spurious (but excluding those making up the polar-cap black-outs) in this second plot was necessary in order to permit a comparison to be made with the magnetic data. Since auroral *B*'s had already been selected on the basis of magnetic activity, an index compiled only from auroral *B*'s could not be expected to give any meaningful results when compared with an index of magnetic activity. The daily sum of *K* indices from Meanook were plotted in the same manner with a 3-division scale: (1)  $K < 20$ , (2)  $20 \leq K \leq 30$ , (3)  $K > 30$ . Figure 7 shows the 11-year plots of these magnetic and black-out daily indices. International 27-day intervals beginning with solar rotation number 1582 have been used with a repetition of the first 6 days at the end of each horizontal line. Part A of the figure gives the magnetic activity as observed at Meanook. The years of high and low activity are immediately apparent in the general configuration of the plot and the long sequences of recurrent magnetic storms (1950-54) which have been investigated by Bhargava and Nagvi (1954) are also clearly evident. It can readily be seen that there is a marked difference in the general appearance of the plot for the years in which activity is declining (1950-54) and increasing (1955-58).

Parts B and C of Fig. 7 show the daily black-out indices for Churchill for the 11-year period. C indicates only the polar-cap black-outs. Here again the years of high and low activity are clearly evident although there seems to be little difference between the periods of increasing and decreasing activity. In part B all black-out occurrences were considered other than those shown in part C. Part B therefore includes auroral black-outs and those black-outs which were believed to be spurious. This section of the diagram exhibits quite a different pattern of variation. There is no clear minimum in the black-out occurrences (part B) around 1954 at the time of minimum magnetic activity. The most striking feature perhaps is the obvious similarity between the general configurations of the magnetic and black-out indices for the period 1950-54 even though part B is made up from all except polar-cap black-out occurrences. In spite of some contamination of the ionosonde data due to low critical frequencies, the 27-day recurrent sequences show up quite clearly in the plots of the black-out indices. In contrast to this it can be seen that parts A and B of the figure for the period 1956-59 are quite dissimilar. During these years the increase in magnetic activity was quite clearly not accompanied by an increase in the occurrence of auroral black-outs. This confirms the conclusion already drawn from Fig. 4. The inclusion in part B of the short periods of auroral black-out from the polar-cap black-outs in part C would not change the general features of these parts of the figure.

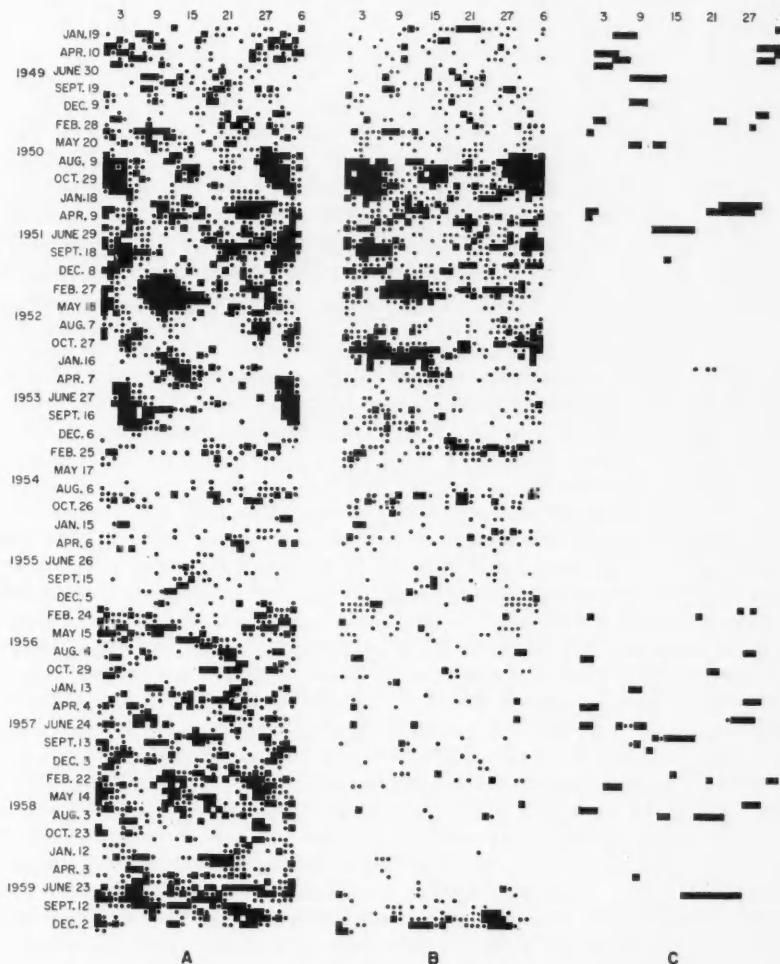


FIG. 7. 27-day recurrence time patterns; (A) magnetic activity at Meanook, (B) all black-outs at Churchill, excluding polar-cap black-outs, (C) polar-cap black-outs at Churchill.

Absorption index:  $\square$  daily sum of hourly  $B$ 's = 0;  $\bullet$  daily sum of hourly  $B$ 's  $< 3$ ;  $\blacksquare$  daily sum of hourly  $B$ 's  $> 3$ .

Magnetic index:  $\square$  daily sum of  $K$  values  $< 20$ ;  $\bullet$  daily sum of  $K$  values  $20 < K < 30$ ;  $\blacksquare$  daily sum of  $K$  values  $> 30$ .

#### (e) Latitude Dependence of Auroral Black-outs

In order to display the latitude dependence of auroral black-outs some of the results from Fig. 4 have been shown in a different way in Fig. 8. In this plot the number of auroral black-outs has been shown as a function of geomagnetic latitude for every other year starting with 1949. It will be seen that

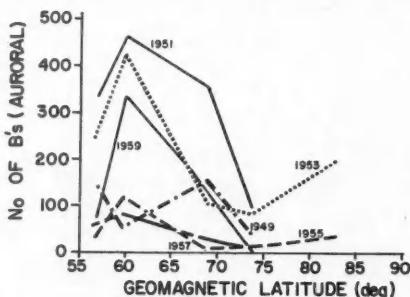


FIG. 8. Yearly totals of auroral  $B$ 's shown as a function of geomagnetic latitude for alternate years from 1949 to 1959.

a maximum in the number of auroral  $B$ 's occurs at Churchill only in 1949. For all other years the peak occurs at Winnipeg. Plots of all  $B$ 's (excluding polar-cap black-outs) produce curves giving a very similar result.

#### DISCUSSION

An earlier statistical study by Agy (1954a), based on h.f. propagation between geomagnetic latitudes 45 and 65 degrees along the 90th meridian, presented evidence for a narrow auroral absorption zone centered about the latitude of maximum occurrence of visual aurora. Cox and Davies examined the occurrence of all  $B$ 's at Resolute Bay, Baker Lake, Churchill, and Winnipeg for the years 1949-1952 and found a pronounced peak in the frequency of occurrence at Churchill. Although a close comparison cannot be made between the present study and the earlier ones it is felt that the results presented here suggest a somewhat different picture of the absorbing regions.

Some appreciation of the latitude variations of the absorption can be obtained from the yearly totals in both Fig. 4 and Fig. 8. It should be remembered, however, that polar-cap black-outs have been excluded from the plots in both figures and the curves in Fig. 8 represent only the numbers of auroral  $B$ 's. The inclusion of the  $B$ 's comprising the polar-cap black-outs would certainly change the shape of some of the histograms and curves but only for the higher-latitude stations. For Resolute Bay, where polar-cap black-outs account for 85% of the black-out occurrences during years when the sunspot number is greater than 100, the number of  $B$ 's would be increased and the outline of the histogram would follow the sunspot curve closely. A somewhat similar condition obtains for Baker Lake. For Churchill the yearly totals of all  $B$ 's do show the effect of the polar-cap black-outs but the general agreement with sunspot variation is poor. For Winnipeg and Ottawa the  $B$ 's during polar-cap black-out events form a small part of the total observed. Totalling all black-out occurrences would therefore present a different picture of the latitude dependence at only the northern stations. South of Baker Lake there would still seem to be an extensive absorbing region. It seems likely that there is a maximum in this region but the present results would not lead one to

expect a 'narrow' absorption zone. From the curves in Fig. 8 one sees that during 1949 the peak did occur at Churchill but for some years there were as many *B*'s recorded at Winnipeg as at Churchill and for other years there were more.

It may be argued that the method of selecting the data is entirely responsible for this somewhat different picture. It is known, however, that the two types of absorption discussed here often occur simultaneously. A specific event of this kind has been described by Reid and Collins from riometer measurements at Churchill. It seems likely that similar conditions occur at Baker Lake and Winnipeg. This simultaneity of the two types cannot be detected in the ionosonde data and therefore the most meaningful way of presenting the synoptic picture of auroral black-outs is to exclude the periods when polar-cap absorption is known to occur.

Some consideration must also be given to the effect of the magnetic activity criterion on the relative frequency of occurrence of *B*'s for the different stations. Fel'dshstein and Kurdina (1958) have shown that the general character of the magnetic field variation is essentially the same over a distance of 180 km along the meridian and over a distance of 300 km in longitude in regions near the auroral zone. It can thus be assumed that the Agincourt *K* index is a good measure of the magnetic disturbance at Ottawa 370 km away. Whitham and Loomer (1956) have shown that there is a high degree of correlation between the *K* indices for Meanook and Baker Lake; these stations are separated by a distance of approximately 1500 km. However, the distance between Meanook and Winnipeg is slightly less than the distance between Meanook and Churchill. This and the latitudes of the stations may cause some weighting of the Winnipeg *B*'s when the Meanook *K* index is used as the criterion. It is felt, however, that this is not an important factor. If one assumes the most unfavorable conditions for Winnipeg, that all the doubtful cases of black-out should be rejected for Winnipeg and none for Churchill, the general picture of the latitude dependence is changed only slightly and then only for a part of the 11-year period.

The decrease in the occurrence of auroral black-outs north of Churchill seems to persist throughout the solar cycle and agrees with the results from earlier studies. It is difficult, however, to account for the black-out occurrences at Resolute Bay which are shown in Fig. 4. It seems possible that a large number of these auroral black-outs at Resolute Bay are short-lived polar-cap black-outs and that the true auroral black-outs may occur very infrequently north of Baker Lake. Although these black-outs occurred frequently on many consecutive days during 1950-52, they were short-duration events which did not meet the criteria laid down for the selection of the polar-cap black-outs; nevertheless, it may be that they are indeed manifestations of very weak polar-cap black-outs. Similar conditions were obtained for Baker Lake and Churchill, to a lesser extent. *K* indices from Resolute Bay were available for only 4 of the 11 years, and so they could not be applied in the selection of auroral black-outs at this station. However, it is unlikely that such a selection could be justified by the arguments presented earlier, since very little is known

about the association of magnetic and ionospheric disturbances at very high latitudes. It is even questionable whether it is meaningful to refer to these black-outs as 'auroral'. Visible aurora occurs very infrequently at Resolute Bay. Several workers have postulated an inner auroral zone and Lassen (1959) has recently described a region of visual aurora at geomagnetic latitudes of approximately 75 to 80 degrees but there is as yet no evidence for associating this aurora with the very high latitude black-outs. Reid (1960) has reported that riometer measurements at Thule ( $\Phi = 87$  degrees) showed a complete absence of auroral absorption during the I.G.Y. These were years of high solar activity. It may be that only during years of declining solar activity are the auroral black-outs detected at very high latitudes. At such times the short-duration absorption events may be due to a combination of the two kinds of black-outs. A meaningful interpretation of the Resolute Bay 'auroral' black-outs must be based on the riometer data now being collected.

Because of the coarseness of the ionosonde measurements little can be said about the extent of the polar-cap absorption region. While no polar-cap black-outs were observed at Ottawa, about half of these events were accompanied by black-outs at Winnipeg before the time of sudden commencement. These were of much shorter duration than at the higher-latitude stations. Since Ottawa and Winnipeg differ in geomagnetic latitude by only 2.3 degrees this result suggests a rather sharp southern boundary for the polar-cap absorbing region and tends to confirm what is already known about the spatial extent of the phenomenon (Reid and Leinbach 1959).

Although much more than 11 years of data is required for the determination of the secular variation of these absorption phenomena an interesting comparison can be made between the variations presented here in the yearly totals of the two kinds of black-out. It is perhaps not surprising that one finds such close agreement between the sunspot numbers and the occurrences of polar-cap black-out since the polar-cap absorption is almost always associated with the occurrence of large solar flares, which, in turn, are associated with large and active sunspot groups. On the other hand the disagreement between the sunspot number curve and the yearly totals of auroral black-outs is quite striking. It is possible that the dependence of the black-out condition on *F*-region critical frequency is contributing to this result. However, the monthly median noon values of  $f_0F_2$  are known to follow the smoothed sunspot number variations very closely. Any significant dependence of the black-out condition on critical frequencies should have produced a marked increase in the occurrence of black-outs at the time of sunspot minimum (1954). This was clearly not the case. It might also be suggested that changes in data-scaling procedures and modifications to equipment during the past 11 years may be factors affecting the number of black-outs reported. Again it is felt that these are minor considerations. Although noise-level measurements are not available from the ionosonde stations, transmitter powers have been kept the same and the sensitivities of the sounders are considered to be reasonably comparable. The fact that all the histograms in Fig. 4 have the same general shape lends weight to the argument that this secular variation of the auroral black-outs is real and has a physical rather than an instrumental cause. It is also felt

that the secular variation of auroral *B*'s as shown in Fig. 4 is not seriously affected by the omission of the auroral black-outs which occurred during polar-cap black-outs. The variation in the yearly totals of all auroral *B*'s at Ottawa (times of polar-cap black-out included) was found to be very similar to that shown by the histogram in Fig. 4.

Since nearly all measures of visible solar activity correlate well with the variations of sunspots it is difficult to suggest an optical feature of the sun which might be associated with the occurrence of the auroral black-outs. In this respect it is interesting to recall the work of Babcock and Babcock (1954). These authors have presented evidence for the existence of both general and localized solar magnetic fields and discussed the importance of these in relation to the emission of corpuscular streams from the sun. In a more recent paper Babcock (1959) has given a further description of the variations in polarity and intensity of the general solar field. When the field was first detected in 1952 it showed a polarity reversed to that of the earth. This was maintained through the years of sunspot minimum until some time between March and July, 1957. At that time the south polar field reversed its sign. The sign of the north polar field remained unchanged until November, 1958, when it rather abruptly became negative. Although there is no evidence to suggest that the two phenomena are related, it is interesting to note that it was just about this time that the auroral black-out occurrences reached a minimum.

It is well known that the solar features during the period of declining spot activity are not the same as during the period of increasing activity. Terrestrial phenomena also exhibit different behavior in different parts of the solar cycle. Large magnetic storms recur at 27-day intervals more frequently in years of decreasing sunspot numbers than in the years when the numbers are increasing. Similar recurrence tendencies have been observed at these times in the occurrence of visual aurora. Such conditions were certainly in evidence during the period 1950 to 1953. The 27-day recurrence pattern which can be seen in the 11-year plot of the daily black-out index, and the maxima about 1950-53 in the histograms in Fig. 4, strongly suggest that auroral black-outs and polar-cap black-outs must be associated with different basic conditions on the sun. It is during the years of decreasing spot activity that the solar *M* regions reach the peak of their activity (Chapman and Bartels 1940). Little seems to be known, however, about these *M* regions although Abetti (1957) has pointed out that such regions may be identified with the filament prominences. The same author has also suggested that the very active coronal regions which have no counterpart disturbances in the photosphere or chromosphere may be connected in some way with the *M* regions. Babcock and Babcock (1954) have suggested that the occasional extended magnetic areas of only one outstanding polarity, which they term unipolar magnetic (UM), may be identified as the *M* regions. Whatever their nature it seems certain that these regions play an important role in the generation of the low-energy particles believed to be responsible for ionospheric disturbances in the vicinity of the auroral zone. It may thus be that the auroral type of absorption is more closely associated with the solar *M* regions than with the flares and sunspots.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the many helpful discussions held with Dr. I. Paghis and Dr. T. R. Hartz of the Radio Physics Laboratory, D.R.T.E.

## REFERENCES

ABETTI, G. 1957. *The sun* (Faber and Faber, London).

AGY, V. 1954a. *J. Geophys. Research*, **59**, 267.

— 1954b. *J. Geophys. Research*, **59**, 499.

BABCOCK, H. D. 1959. *Astrophys. J.* **130**, 364.

BABCOCK, H. W. and BABCOCK, H. D. 1954. *Astrophys. J.* **121**, 349.

BAILEY, D. K. 1959. *Proc. I.R.E.* **47**, 255.

— 1957. *J. Geophys. Research*, **62**, 431.

BHARGAVA, B. N. and NAGVI, A. 1954. *Nature*, **173**, 498.

CHAPMAN, S. and BARTELS, J. 1940. *Geomagnetism* (Clarendon Press, Oxford).

COX, J. W. and DAVIES, K. 1954. *Can. J. Phys.* **32**, 743.

DAVIES, K. and HAGG, E. L. 1955. *J. Atmospheric and Terrest. Phys.* **6**, 18.

FEL'DSHTEIN, YA. I. and KURDINA, YE. I. 1958. *Problemy Arktiki*, **3**, 53.

HAKURA, Y., TAKENOSHITA, Y., and OTSUKI, T. 1958. *Rept. Ionosphere Research Japan*, **XII**, 459.

HEPPNER, J. P., BYRNE, E. C., and BELON, A. E. 1952. *J. Geophys. Research*, **57**, 121.

LASSEN, K. 1959. *Nature*, **184**, 1375.

LINDQUIST, R. 1951. *Trans. Chalmers Univ. Technol. Gothenburg*, 103.

LITTLE, C. G. and LEINBACH, H. 1958. *Proc. I.R.E.* **46**, 334.

LITTLE, C. G., RAYTON, W. M., and ROOF, R. B. 1956. *Proc. I.R.E.* **44**, 992.

MEEK, J. H. 1952. *J. Geophys. Research*, **57**, 177.

— 1953. *J. Geophys. Research*, **48**, 445.

— 1957. The Canadian Program for the I.G.Y. Association Committee on Geodesy and Geophysics, National Research Council, Ottawa.

REID, G. C. 1960. Private communication.

REID, G. C. and COLLINS, C. 1959. *J. Atmospheric and Terrest. Phys.* **14**, 63.

REID, G. C. and LEINBACH, H. 1959. *J. Geophys. Research*, **64**, 1801.

WELLS, H. W. 1947. *Terrestrial Magnetism and Atmospheric Elec.* **52**, 315.

WHITHAM, K. and LOOMER, E. I. 1956. *Tellus*, **8**, 276.

## NUCLEAR ORIENTATION IN ANTIFERROMAGNETIC SINGLE CRYSTALS<sup>1</sup>

J. M. DANIELS,<sup>2</sup> J. C. GILES, AND M. A. R. LEBLANC<sup>3</sup>

### ABSTRACT

Mn54 and Co60 have been successfully oriented in five antiferromagnetic single crystals ( $MnCl_2 \cdot 4H_2O$ ,  $MnBr_2 \cdot 4H_2O$ ,  $CoCl_2 \cdot 6H_2O$ ,  $Co(NH_4)_2(SO_4)_2 \cdot 6H_2O$ , and  $MnSiF_6 \cdot 6H_2O$ ) and the orientation was detected by the anisotropy of the emitted  $\gamma$  rays. Only in the case of Co60 in  $MnBr_2 \cdot 4H_2O$  was no  $\gamma$ -ray anisotropy seen. It is concluded that antiferromagnetism can be used as a means of producing nuclear orientation. Attempts to orient Br82 and I131 in the manganese halides by superexchange were unsuccessful.

### 1. INTRODUCTION

One of the most successful methods of orienting nuclei is by the use of the magnetic hyperfine structure. This makes use of the fact that the field produced by an unfilled electron shell at the nucleus of an atom is very large, of the order of  $10^6$  gauss for an unfilled  $3d$  shell. The nuclei have different energies for different orientations in this field, and at temperatures, below about  $0.1^\circ K$ , the lower energy levels are significantly preferentially populated. If any means can be found to orient these fields parallel, this preferential population of nuclear orientation levels corresponds to parallel orientation of the nuclei. The internal atomic fields can be oriented if the electron spins in the unfilled shell can be so oriented. The foregoing is an oversimplification of the situation, but it brings out the essential features.

One method of orienting the electron spins parallel, is to make use of the interaction which produces antiferromagnetism. In antiferromagnetic substances below the Néel temperature, the electron spins are (in the simplest case) ordered alternately parallel and antiparallel to a particular direction in the crystal. If this direction is known, all that is needed is to cool a single crystal to a temperature below its Néel point and below about  $0.1^\circ K$ , and the nuclei will become aligned parallel to this direction of spontaneous magnetization of the sublattices. Such a mechanism for producing nuclear alignment appears to have been first suggested by Daunt (1951) and Gorter (1951). The original attempts to align nuclei in antiferromagnetic substances gave negative results (Poppe 1954), and the possibilities of this method appeared poor when it was observed that interactions in paramagnetic salts tended to reduce the nuclear alignment, especially below the Néel point (Grace *et al.* 1954). Interest in this method was revived by measurements of the specific heat of  $MnF_2$  below  $1^\circ K$  (Cooke 1957; Cooke and Edmonds 1958). A large contribution to the specific heat was observed which could be reasonably ascribed only to the hyperfine structure of the manganese ions. The first

<sup>1</sup>Manuscript received March 7, 1960.

Contribution from the Department of Physics, University of British Columbia, Vancouver 8, B.C.

<sup>2</sup>On leave of absence at Instituto de Fisica, San Carlos de Bariloche, (R.N.), Argentina.

<sup>3</sup>Now at Department of Physics, Stanford University, Palo Alto, California.

successful alignment of nuclei in an antiferromagnetic crystal was reported by Daniels and LeBlanc (1958). We now wish to present the results of a survey investigation, in which we have established the feasibility of this method for nuclear alignment.

Before presenting these results, let us review some of the questions which have arisen in connection with this method. First there is the problem of cooling the nuclei. This can be divided into two parts, the cooling of the crystal itself, and the relaxation processes whereby the nuclei cool to the temperature of the rest of the crystal. Antiferromagnetic substances cannot be cooled by adiabatic demagnetization, since the temperature dependence of the susceptibility has the wrong sign. The crystal must therefore be cooled by contact with a heat sink at a low temperature, and in our experiments we used chromium potassium alum cooled by adiabatic demagnetization. Thermal conductivities in this temperature range are poor, and ingenuity must be used to find a suitable experimental arrangement which provides efficient heat transfer. A brief review of some of the methods which we used is given in Daniels and LeBlanc (1959b); the final arrangement which we used is that given in Daniels and LeBlanc (1959a). Thermal equilibrium was attained in about 20 minutes. Nuclear spin-lattice relaxation times are always long at these low temperatures, the direct interaction with the lattice produces relaxation times of the order of years. The most efficient coupling with the lattice is through the unpaired electron spins, and this presents certain interesting features. In the antiferromagnetic state, the electron spins are closely coupled to one another, and the nuclei do not relax by giving up their energy to a single electron, but rather to the whole lattice of electron spins. This is most conveniently pictured as nuclei flipped by spin waves, each flip either creating or changing the energy of a spin wave. Now spin waves have a rest energy which is of the order of magnitude of  $kT_A$  (where  $T_A$  is the anisotropy temperature; about half the Néel temperature in the salts we used), hence the number of spin waves at temperature  $T$  will be proportional to  $e^{-T_A/T}$  (among other things). Thus, at low temperatures, very few spin waves will be excited, and the relaxation time can be expected to contain a factor  $e^{T_A/T}$  which increases rapidly as the temperature is diminished. Theories of this relaxation process have been given by Moriya (1956) and by Bloom and van Kranendonk (1956), and such long relaxation times have been observed experimentally by Cooke and Edmonds (1958) in  $\text{MnF}_2$ . Hence there is an advantage in using a salt with a low Néel temperature. One of the predictions of Moriya's theory is that nuclei such as Mn whose hyperfine coupling is isotropic (of the form  $A \mathbf{S} \cdot \mathbf{I}$ ), should relax more slowly than others, such as Co, whose hyperfine coupling is anisotropic and there is the possibility that the principal axes of the h.f.s. do not coincide with the principal axes of the antiferromagnetic anisotropy. These experiments gave us an opportunity to test this prediction.

Further, it is not obvious that salts with a low Néel temperature should follow the classic pattern of antiferromagnetism, because crystalline field interactions are stronger than the antiferromagnetic coupling, and in some cases even the hyperfine coupling is stronger than the antiferromagnetic

interaction (Daniels 1959). The predominance of the crystalline field interaction could lead to the appearance of more than two sublattices, whose directions of spontaneous magnetization are dictated by the crystalline field rather than by the magnetic interaction.

One reason why it was once thought that this method of nuclear alignment would not work, is that the basic state of an antiferromagnetic is degenerate; reversing the directions of all the electron spins leads to a state of the same energy. The ground state should therefore be a mixture of these two basic states in approximately equal proportions, the magnetization of each sublattice changing direction from time to time, so that the average field produced at the nuclei would be zero. Although the true ground state of an antiferromagnetic is considerably more complicated than this, this is a sufficiently good approximation for many purposes, and is borne out by neutron diffraction experiments. The appearance of a hyperfine structure in  $MnF_2$  indicates that the two sublattices do not exchange identities so rapidly that the nuclei cannot follow; but there remains the possibility that the field at a nucleus might be proportional to the spontaneous magnetization of a single sublattice, and hence the hyperfine splitting might vanish, or go through a minimum, at the Néel temperature. This would imply that the anisotropy of emission of  $\gamma$  rays from nuclei oriented in these salts would go through a minimum at this temperature. Unfortunately, we are able to test this supposition only with salts whose Néel temperature is less than about  $0.1^\circ K$ , for which the hyperfine interaction is of the same order of magnitude as the antiferromagnetic interaction.

## 2. RESULTS

The technique used for measuring anisotropy of  $\gamma$ -ray emission is now well known, and can be found described in many publications. The following, for example, give all the details pertinent to this series of experiments: Daniels and LeBlanc (1959a), Bleaney *et al.* (1954). The  $\gamma$  rays were counted along two directions at right angles; let us call these directions  $\alpha$  and  $\beta$ . The direction  $\alpha$  is the direction along which the nuclear spin axes are expected to be aligned, and the anisotropy  $\epsilon$  is defined as

$$\epsilon = \{I(\beta) - I(\alpha)\}/I(\beta),$$

where  $I(\alpha)$  and  $I(\beta)$  are the normalized counting rates in the directions  $\alpha$  and  $\beta$  respectively. We have measured the anisotropy of  $\gamma$ -ray emission from  $Co_{60}$  and  $Mn_{54}$  in five salts listed in Table I; and a brief indication of the results is

TABLE I  
Salts used and their Néel temperatures ( $^\circ K$ )

$MnCl_2 \cdot 4H_2O$	1.6
$MnBr_2 \cdot 4H_2O$	2.2
$CoCl_2 \cdot 6H_2O$	3.0
$Co(NH_4)_2(SO_4)_2 \cdot 6H_2O$	0.085
$MnSiF_6 \cdot 6H_2O$	0.1

given in Tables II and III. The results of a typical experiment are shown in Fig. 1 which depicts the simultaneous measurements of  $\gamma$ -ray anisotropy of Co60 and Mn54 from a  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  crystal containing a few microcuries of

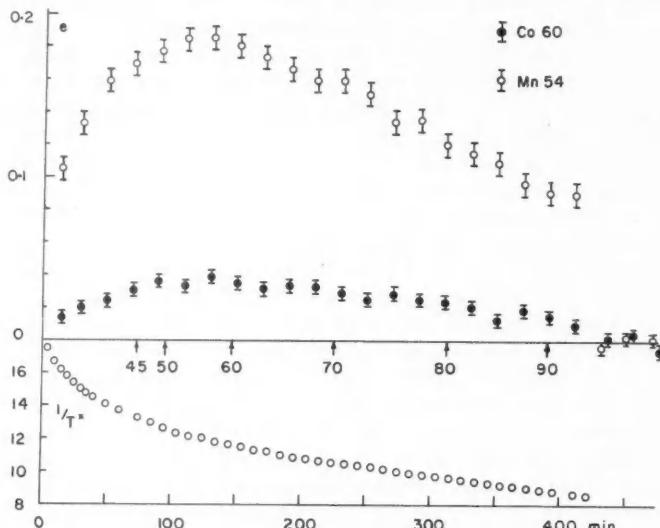


FIG. 1. The anisotropy,  $\epsilon$ , of  $\gamma$  radiation from Mn54 and Co60 in single crystal of  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ , and the magnetic susceptibility ( $1/T^*$ ) of the heat sink, as a function of time from demagnetization. The figures below the arrows are the absolute temperature of the heat sink at that time in millidegrees. After 425 minutes, the whole sample was warmed to the temperature of the helium bath.

TABLE II  
Anisotropies seen from Co60

Salt	Max $\epsilon$	$T$ for max $\epsilon$ ( $^{\circ}\text{K}$ )	0.07 $^{\circ}\text{K}$	0.08 $^{\circ}\text{K}$	0.09 $^{\circ}\text{K}$	0.10 $^{\circ}\text{K}$
$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	.007	0.1	.....	All < 0.005	.....	0.007
$\text{MnBr}_2 \cdot 4\text{H}_2\text{O}$	None					
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	0.04	0.055	0.04	0.025	0.015	
$\text{Co}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$	0.06	0.045	0.025			
$\text{MnSiF}_6 \cdot 6\text{H}_2\text{O}$	0.19	0.05	0.11	0.08	0.06	0.035

TABLE III  
Anisotropies seen from Mn54

Salt	Max $\epsilon$	$T$ for max $\epsilon$ ( $^{\circ}\text{K}$ )	0.07 $^{\circ}\text{K}$	0.08 $^{\circ}\text{K}$	0.09 $^{\circ}\text{K}$	0.10 $^{\circ}\text{K}$	0.11 $^{\circ}\text{K}$	0.12 $^{\circ}\text{K}$
$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	0.09	0.09			0.09	0.085	0.075	0.06
$\text{MnBr}_2 \cdot 4\text{H}_2\text{O}$	0.055	0.09			0.055	0.05		
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	0.18	0.055	0.155	0.125	0.095	0.07		
$\text{Co}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$	0.18	0.055		0.065	0.045	0.035		
$\text{MnSiF}_6 \cdot 6\text{H}_2\text{O}$	0.20	0.055	0.16	0.12	0.09	0.07	0.045	0.025

both isotopes. The manganese and cobalt  $\gamma$  rays were separated by a discriminator and a single channel kicksorter, and counted on separate scalers, and a correction was applied for the cobalt background appearing in the manganese channel. The anisotropies quoted in Tables II and III were, however, obtained from measurements on salts containing only one radioactive isotope.

In all the experiments, the maximum anisotropy was seen some time after demagnetization, this time being usually about 2 hours in the earlier experiments, and about 20 minutes in the later ones, as the technique was improved. As was explained in the Introduction, this delay could be due either to bad heat conduction, or to a long nuclear spin-lattice relaxation time. Further, the theories of relaxation predict that the nuclear spin-lattice relaxation time for manganese should be larger than that of cobalt by a factor of several powers of 10. Figure 1 is typical of most of the results obtained from crystals containing both Co60 and Mn54; the anisotropies of  $\gamma$  rays from both isotopes reach a maximum at the same time, indicating that in these salts at least, the time needed to cool the nuclei to about 0.05° K depends on heat conduction rather than on nuclear spin-lattice relaxation. We now review the results of measurements on the various salts in turn.

#### $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ and $\text{MnBr}_2 \cdot 4\text{H}_2\text{O}$

These salts are isomorphous, and the most complete account of their magnetic properties at low temperatures is given by Gijsman *et al.* (1959). They crystallize in the monoclinic system, and an account of their external morphology is given in Groth (1906). The angle between the *a* and *c* axes is 99°25' for the chloride and 99°6' for the bromide, the *b*–*c* plane is well developed and the *b* and *c* axes are easily identified. According to Gijsman *et al.* (1959) the preferred orientation in the antiferromagnetic state is the *c*-axis in both salts. The next preferred orientation is the *b*-axis in the chloride, and it is assumed to be the same for the bromide, although this has not been directly observed. Susceptibility measurements show that the magnetic moment of the manganese ion is isotropic. There are no paramagnetic resonance data available, nothing is known about the crystal field splittings of the manganese ions, and the crystal structure is unknown. The Néel temperatures are 1.6° K for the chloride and 2.2° K for the bromide. The directions along which the  $\gamma$  rays were counted,  $\alpha$  and  $\beta$  above, were the *c*-axis and the *b*-axis respectively.

At low temperatures, an anisotropy is developed in the  $\gamma$ -ray emission from the Mn54, which is consistent with the hypothesis that the nuclei are aligned along the crystallographic *c*-axis. In both salts, the anisotropy reaches a maximum about six hours after demagnetization, even with the latest techniques for cooling. We can therefore ascribe this delay to a long nuclear spin-lattice relaxation time, since we know that these cooling techniques are sufficiently good to cool other salts in 1/10th of this time. Owing to this long relaxation time, it was impossible to obtain reliable values of the  $\gamma$ -ray anisotropy as a function of temperature. The maximum anisotropies seen are 0.085 in  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  and 0.055 in  $\text{MnBr}_2 \cdot 4\text{H}_2\text{O}$  and we estimate the

temperature to be approximately  $0.1^\circ$  K at this maximum. We did not see a true maximum anisotropy in  $\text{MnBr}_2 \cdot 4\text{H}_2\text{O}$ ; the anisotropy was still growing 10 hours after demagnetization in one experiment.

When  $\text{Co}_{60}$  was introduced into these salts as an impurity, presumably replacing manganese, no anisotropy was seen in the bromide, and only about 0.007 in the chloride. This latter figure is hardly larger than the statistical error.

Nothing is known about the crystal structure, nor about the magnetic properties of manganese or cobalt in this lattice. In the absence of this information we are left to conjecture, and the following seems a plausible explanation of the small anisotropy. The magnitude of the crystal field splitting in a  $\text{Mn}^{++}$  ion is usually about  $0.2^\circ$  K. This is less than the Néel temperature, hence the co-operative interaction dominates, and the manganese spins are aligned along the *c*-axis under the influence of this interaction, the crystal field effects being feeble and producing only a small distortion of this array. The crystal field effects on a typical  $\text{Co}^{++}$  ion are very much stronger, and will probably dominate the situation. The cobalt spins will be constrained by the crystal field to point along certain (at present unknown) axes in the crystal, and the magnetic interaction will be able to rotate these spins only a few degrees from these directions. The nuclei of  $\text{Co}_{60}$  are thus oriented along some unknown axes, and it is quite fortuitous if any anisotropy of  $\gamma$ -ray emission is seen from  $\text{Co}_{60}$  using the directions of observation which we chose. A similar situation to this is known to occur in the Tutton salts (see the relevant section below, and also Miedema *et al.* 1959).

We thought that perhaps the nuclear spin-lattice relaxation time of the manganese could be shortened by applying an external magnetic field at an angle to the *c*-axis. This could conceivably rotate the magnetization of the electron spins, so that their directions of magnetization might no longer coincide with the principal axes of the spin-Hamiltonian of the manganese ions. Fields of up to 1000 gauss were applied for periods of up to 1 hour, at various times after demagnetization, at  $45^\circ$  to the *c*-axis. However, measurements of the anisotropy after these fields were removed showed that they had no influence on the rate of growth of anisotropy.

Following the experience of Cooke and Edmonds (1958), who succeeded in shortening the nuclear spin-lattice relaxation time of Mn in  $\text{MnF}_2$  by irradiating their crystal with X rays (which presumably produced magnetic radiation-damage centers), it was hoped that a similar procedure might have a similar effect here. Accordingly, an assembly of chrome-alum pills and a crystal of  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  was irradiated as a unit, so as not to disturb the geometrical arrangement (which determines the rate of heat conduction). The X-ray beam was collimated to strike only the manganese chloride crystal. Irradiations were carried out at room temperature for up to 40 hours (copper target, 40 kv, 15 ma) but no decrease was observed in the time needed for the appearance and growth of the anisotropy.

#### $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$

This salt crystallizes in the monoclinic system, and its external morphology has been described by Groth (1906). The angle between the *a* and *c* axes is

122°19'. The crystal cleaves easily in the (001) plane, this fact is useful in identifying the axes. The complete crystal structure has recently been determined by X rays (Mizumo *et al.* 1959). There are two cobalt ions in the unit cell which are magnetically equivalent. Each cobalt ion is surrounded by four water molecules in a plane, and two chloride ions in a direction perpendicular to this plane, these six forming an octohedron. The Cl—Co—Cl directions lie in the  $a-c$  plane and make an angle of 103°45' with the  $c$ -axis lying in the obtuse angle between  $a$  and  $c$ .

The most up-to-date data on the magnetic properties of this salt are given by Flippin and Friedberg (1960), who state that the preferred direction of magnetization is the crystallographic  $c$ -axis. Paramagnetic resonance measurements by Date (1959) indicate that the  $g$  values along the  $a'$ ,  $b$ , and  $c$  axes are 2.9, 5.0, and 4.0 respectively ( $a'$  is an axis perpendicular to  $b$  and  $c$ ). These vary slightly from the values obtained from susceptibility measurements, but indicate that, to a first approximation, the cobalt ion has tetragonal symmetry, with  $g_{\parallel} \simeq 2.7$  and  $g_{\perp} \simeq 5$ . It is unusual to find a cobalt ion with  $g_{\perp} > g_{\parallel}$ . Nothing is known about the hyperfine coupling constants, and nothing is known about the magnetic properties of manganese in this salt, as an impurity replacing cobalt.

When we started this work, we were unaware of any of the above data except what was given in Groth, and we determined the Néel temperature, and that the preferred direction of magnetization was approximately the  $c$ -axis, independently of the Japanese investigations. The detailed data on the crystal structure and the magnetic properties came to our attention only after the completion of the alignment experiments. Accordingly, we chose the  $c$  and  $b$  axes as the directions  $\alpha$  and  $\beta$  for the counting. The results of a typical experiment are shown in Fig. 1. We may consider that the crystal field acting on the cobalt ion constrains the "effective spin" to lie in, or close to, a plane perpendicular to the Cl—Co—Cl direction, and that the magnetic interaction removes what freedom is left, constraining the spin to lie in the  $c$ -axis. In the absence of any data on the hyperfine interaction terms in the spin-Hamiltonian, it is impossible to say what would be the angular distributions of  $\gamma$  rays, but the anisotropy observed, although small, is not inconsistent with a tendency of the nuclei to align parallel to the  $c$ -axis. If we assume that, in the case of manganese, the magnetic interaction dominates the crystal field effects, then we should expect the manganese nuclei to become aligned along the  $c$ -axis. The large anisotropy observed (maximum 0.18 at 0.055° K) lends support to this view. Without knowing the spin-Hamiltonian of manganese in this salt, we cannot draw any further conclusions. The short cooling time, and the simultaneous development of the anisotropies in Co60 and Mn54, seem to indicate that the delay in cooling is due to bad heat conduction, rather than to nuclear spin-lattice relaxation.

Recent experiments have been made on manganese in  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  in this laboratory, using three detectors aligned along the crystallographic  $a$ ,  $b$ , and  $c$  axes of the crystal. Preliminary evidence favors the view that there is not a single axis of alignment, but the data is insufficient at present to indicate the orientation of possible multiple alignment axes within the crystal.

$\text{Co}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$

This salt is one of a series which has been extensively studied from many aspects. The external morphology has been described by Tutton (1905, 1913, 1916) and the crystal structure as determined by X rays has been published by Hoffmann (1931). Paramagnetic resonance data for  $\text{Co}^{++}$  and  $\text{Mn}^{++}$  in this lattice are given in Bowers and Owen (1955). The Néel temperature,  $0.085^\circ \text{ K}$ , was determined by Garrett (1951), and a thorough investigation of the antiferromagnetic phase of this salt has been reported by Miedema *et al.* (1959). Although the crystallographic properties of these salts are well known to workers in this field, we shall give a brief outline of them.

The crystals are monoclinic, and the angle between the  $a$  and  $c$  axes is about  $105^\circ$  in all the salts of the series. The principal axes of magnetic susceptibility are denoted by  $K_1$ ,  $K_2$ , and  $K_3$ ;  $K_3$  coincides with the crystallographic  $b$ -axis, and  $K_1$  and  $K_2$  lie in the  $a-c$  plane,  $K_1$  making an angle with the  $c$ -axis. In the cobalt salt,  $K_1$  is the axis of greatest susceptibility,  $K_2$  the axis of smallest susceptibility. There are two ions in the unit cell, each surrounded by six water molecules arranged in an octahedron, distorted from regular shape by elongation along one of its diagonals. This diagonal is known as the tetragonal axis for that ion. The two tetragonal axes,  $T_1$  and  $T_2$ , lie in the  $b-K_1$  (or  $K_3-K_1$ ) plane, and make an angle  $\alpha$  with the  $K_1$ -axis. This geometry is shown in Fig. 2. Nuclear alignment in a paramagnetic Tutton salt has shown

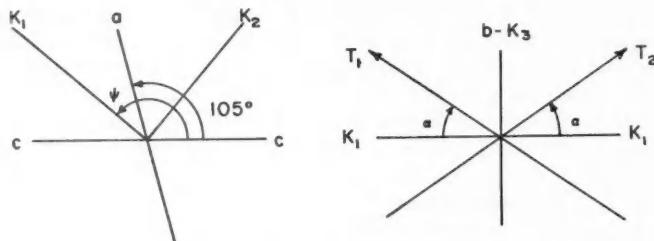


FIG. 2. The crystal axes of  $\text{Co}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  and isomorphous salts. For explanation see text.

that the nuclei of each ion align along the appropriate tetragonal axis (Bleaney *et al.* 1954). The experiments of Miedema *et al.* (1959) show that below the Néel temperature, there are two sublattices, and that these correspond to the two different ions in the unit cell. The electron spins are aligned along the corresponding tetragonal axes in the direction shown by the arrows in Fig. 2, so that there is a net magnetic moment in the direction of the  $b$ -axis. Above the Néel temperature,  $\alpha = 34^\circ$ ; below the Néel temperature, the magnetic interaction, which tries to produce an antiferromagnetic state, rotates the direction of the electron spin axes towards the  $K_1$ -axis, so that  $\alpha = 24^\circ$ . This has been expressed by the statement that "cobalt ammonium sulphate shows ferromagnetic behavior in the direction of the  $b$ -axis, and antiferromagnetic behavior in the  $a-c$  plane".

The angle  $\psi$  ( $c-K_1$ ) for cobalt ammonium sulphate is  $137^\circ$ . For our first experiments we chose  $K_1$  and  $K_2$  as the directions  $\alpha$  and  $\beta$  for our counters. An anisotropy of the  $\gamma$  emission of Co60 was seen, and these results have already been published (Daniels and LeBlanc 1958). The anisotropy of  $\gamma$  emission observed was very closely that which had been observed in a paramagnetic Tutton salt. This led us to the hypothesis that the crystalline field dominates the situation, and constrains the electron spin of the cobalt ion (and hence the nucleus) to point along the direction of the tetragonal axis, and that the magnetic interaction would determine in which sense the spin points along that tetragonal axis, but would not be strong enough to rotate the direction of spin alignment far from this axis. This hypothesis appears to have been confirmed by the work of Miedema *et al.* (1959). We carried out a series of experiments in which the anisotropy of the Co60  $\gamma$  rays was measured as the salt warmed through its Néel temperature. No discontinuity or minimum was observed at or near the transition temperature. This indicates that, contrary to the supposition mentioned in the Introduction to this paper, the sublattices are not rapidly exchanging identities, and is in accord with the observations of Miedema *et al.* (1959) on the magnetic state of the salt below the Néel temperature. As was mentioned in Daniels and LeBlanc (1958), this experiment is of historical interest, since it was tried unsuccessfully by the Leiden group when they were trying to establish a method of nuclear alignment (Poppema 1954). Since the Néel temperature is so low, considerable cooling will occur in the crystal itself on demagnetization, and a measurable anisotropy ought to have been seen immediately after demagnetization. In order to test this point, we demagnetized a crystal of cobalt ammonium sulphate held between two pressed pills of cobalt ammonium sulphate. The magnetizing field was parallel to the  $K_1$ -axis of the crystal, hence the crystal cooled to a lower temperature than the pressed pills after demagnetization. An anisotropy of 0.035 was seen immediately after demagnetization, as was expected. A similar experiment performed on  $\text{MnSiF}_6 \cdot 6\text{H}_2\text{O}$  gave a similar result.

There is at present a controversy about the direction of the  $K_1$ -axis for manganese ammonium sulphate, the position given by Bleaney and Ingram (1951) differing by  $90^\circ$  from that given by Miedema *et al.* (1959). The position of the  $K_1$ -axis for manganese ions in cobalt ammonium sulphate is unknown, but an idea of where it should be can be obtained from its position in other salts of the series. In general, the axes are more or less in the same place in different members of the series, although variations of up to  $20^\circ$  are not uncommon (see, for example, Bowers and Owen 1955). According to Miedema *et al.* (1959), the value of  $\psi$  for manganese ammonium sulphate is about  $150^\circ$ , this is close enough to the value  $137^\circ$  for cobalt ammonium sulphate that, in view of the variation from one salt to another, we can ignore the difference of  $13^\circ$ . In a crystal of cobalt ammonium sulphate containing both Co60 and Mn54, the anisotropy of the  $\gamma$  rays from both isotopes has the same sign, as though the Mn54 nuclei were aligned along the  $K_1$ -axis of the cobalt ions. Of course, this observation does not necessarily mean that the nuclei are aligned along this  $K_1$ -axis; there are many other possibilities, since the hyperfine

interaction, the crystal field interaction, and the magnetic interaction in  $Mn^{++}$  are all of the same magnitude. It should be possible to tell from the temperature dependence of the anisotropy whether the alignment is principally due to the crystal field splitting or to the magnetic interaction, being as  $1/T^3$  in the first case and as  $1/T^4$  in the second at high temperatures. We do not consider our measurement sufficiently accurate to be able to decide what power of  $T$  is followed.

#### $MnSiF_6 \cdot 6H_2O$

This crystal is much simpler in structure than any of the others used. The external morphology is given in Groth (1906). The crystals grow in hexagonal prisms with pointed ends, the axis of the prism being the trigonal axis of the crystal. The crystal structure is known from X-ray measurements (Pauling 1930). There is one ion per unit cell, and these ions are arranged in a rhombohedral lattice, and their immediate surroundings have the trigonal symmetry of this lattice. This salt is isomorphous with  $ZnSiF_6 \cdot 6H_2O$ , and paramagnetic resonance data for  $Co^{++}$  and  $Mn^{++}$  as impurities replacing  $Zn$  in  $ZnSiF_6 \cdot 6H_2O$  are given in Bowers and Owen (1955). According to these data, both the Mn and the Co nuclei should be aligned along the trigonal axis of the crystal at low temperatures, hence we chose this axis as our direction for the  $\gamma$ -ray counting. The Néel temperature is reported to be  $0.1^\circ K$  (Ohtsubo *et al.* 1958).

The  $\gamma$ -ray anisotropies seen in this salt were relatively large—0.20 for  $Mn^{54}$  and 0.19 for  $Co^{60}$ , both at  $0.050^\circ K$ . There appears to be no doubt that the nuclei are aligned along the trigonal axis of the crystal. The anisotropies were observed as the salt warmed through its Néel temperature, and the graph of anisotropy vs. temperature was smooth and continuous through this point, indicating that the sublattices do not exchange identity rapidly in the anti-ferromagnetic state. This is a somewhat different case from cobalt ammonium sulphate, where the two sublattices are geometrically different as well as magnetically different. The  $\gamma$ -ray anisotropy appears immediately after demagnetization due to the cooling of the crystal by adiabatic demagnetization. Although the anisotropies fit the general picture qualitatively, they do not agree quantitatively with values calculated from the spin-Hamiltonians given in Bowers and Owen (1955). In the case of manganese, the discrepancy is not too serious, since the published value of the crystal field splitting parameter " $D$ " was obtained from measurements on a very dilute sample, and it is known that in many salts the constants of the spin-Hamiltonian change appreciably with concentration. The discrepancy in the case of  $Co^{60}$  is more serious, since the temperature dependence of the anisotropy observed is as  $1/T^3$ , not the expected  $1/T^2$ . The very large anisotropies seen indicate that the hyperfine splitting in the cobalt ion must be some three times as large as in the dilute salt. The quantitative interpretation of the anisotropies observed in this salt will be published separately.

#### *Experiments with $Br^{82}$ and $I^{131}$*

The nature of the magnetic interaction is not understood in detail in hydrated salts, and there is evidence that it may be due to some superexchange pheno-

menon which involves an atom lying between two of the magnetic ions. In some cases a hyperfine splitting has been observed due to the coupling of a nucleus of one of the neighboring atoms with the electrons of the magnetic ion. Examples of this are the hyperfine splitting of F19 in  $MnF_2$  (Baker and Hayes 1957; Shulman and Jaccarino 1958), and of Cl in  $K_2IrCl_6$  (Griffiths and Owen 1954; Griffiths *et al.* 1954). We thought that there might also be a hyperfine splitting of the Br nuclei in  $MnBr_2 \cdot 4H_2O$  which could lead to a nuclear orientation of the Br nuclei, and although there existed no data whatever relevant to this case, we nevertheless decided it would be worth while trying the experiment.

In the experiment with Br82, a solution of  $MnBr_2$  was irradiated in a pile and crystals were grown from this solution. The  $\gamma$  rays of energy 1.45, 1.30, 1.03, 0.77, and 0.56 Mev were counted separately. No anisotropy was seen in any of these  $\gamma$  rays. The lowest temperature attained was about 0.06° K, and this is probably much too high to cause any preferential population of the bromine hyperfine levels, if such a hyperfine splitting exists.

We attempted to incorporate I131 in  $MnCl_2 \cdot 4H_2O$  and  $MnBr_2 \cdot 4H_2O$  to replace the halogen. Carrier free iodine as iodide in solution was added to a solution of the manganese salts and crystals were grown from this mixture. When the iodide solution was added to the manganese solution, a small quantity of fine precipitate was formed which appeared to contain a large fraction of the iodine activity. Although we established that the iodine activity is distributed uniformly throughout the crystals grown from this solution we cannot be sure that the iodine replaced the other halogen substitutionally. The 368-kev  $\gamma$  ray was counted, and no anisotropy was observed; the lowest temperature attained was probably 0.06° K. The same criticism can be applied to these experiments as to the ones with Br82.

### 3. CONCLUSIONS

This work was undertaken to investigate the feasibility of using the antiferromagnetic interaction as a means of orienting nuclei, as proposed by Daunt (1951) and Gorter (1951). We can draw the conclusion from this series of experiments, that the mechanism proposed is a valid one, and that suspicions that antiferromagnetism, *per se*, might be a hindrance to the achievement of nuclear orientation are largely unfounded. When the question is asked, "How will the nuclei be aligned in a given antiferromagnetic crystal?", it must be borne in mind that antiferromagnetism is only one of the many factors which must be taken into consideration, others being the nature of the crystal field interaction, and the hyperfine coupling. The answer will depend on the interplay of all these factors, and no generalizations can be made here. We have, in this series of experiments, examples of many different situations. For manganese in manganese fluosilicate, for example, the antiferromagnetic interaction aids the crystal field in producing a nuclear alignment; while in the case of cobalt in cobalt ammonium sulphate, the crystal field plays the most important part, and the antiferromagnetic interaction is of little consequence. Again, in manganese bromide, we find that the crystal field effect

on manganese is much less than the effect of the antiferromagnetic interaction, and the manganese nuclei seem to align along the preferred directions of magnetization; whereas for cobalt, the crystal field and the magnetic interaction compete to orient the nuclei along different directions.

One unfortunate effect is seen in these experiments, the possibility that antiferromagnetism may in some cases so lengthen the nuclear spin-lattice relaxation times as to make some experiments impossible. From the strictly practical point of view, as well as from the desire to understand these processes better, the question of relaxation times merits further study.

#### ACKNOWLEDGMENTS

We acknowledge gratefully the support of N.R.C. and Atomic Energy of Canada Ltd. for the funds which made this work possible. Two of the authors are indebted to N.R.C. for the award of a postdoctorate fellowship (J.C.G.) and a studentship (M.A.R.L.) during the tenure of which this work was carried out. Our thanks are due to the Department of Mining and Metallurgy at U.B.C. for assistance with X-ray irradiations.

#### REFERENCES

BAKER, J. M. and HAYES, W. 1957. *Phys. Rev.* **106**, 603.  
 BLEANEY, B. and INGRAM, D. J. E. 1951. *Proc. Roy. Soc. (London)*, A, **205**, 336.  
 BLEANEY, B., DANIELS, J. M., GRACE, M. A., HALBAN, H., KURTI, N., ROBINSON, F. N. H., and SIMON, F. E. 1954. *Proc. Roy. Soc. (London)*, A, **221**, 170.  
 BLOOM, M. and VAN KRANENDONK, J. 1956. *Physica*, **22**, 545.  
 BOWERS, K. D. and OWEN, J. 1955. *Repts. Progr. in Phys.* **18**, 304.  
 COOKE, A. H. 1957. *Proceedings of an International Conference on Low Temperature Physics* at Madison, Wisconsin (University of Wisconsin Press), p. 528.  
 COOKE, A. H. and EDMONDS, D. 1958. *Proc. Phys. Soc. A*, **71**, 517.  
 DANIELS, J. M. 1959. *Can. J. Phys.* **37**, 75.  
 DANIELS, J. M. and LEBLANC, M. A. R. 1958. *Can. J. Phys.* **36**, 638.  
 ——— 1959a. *Can. J. Phys.* **37**, 82.  
 ——— 1959b. *Can. J. Phys.* **37**, 1321.  
 DATE, M. 1959. *J. Phys. Soc. Japan*, **14**, 1244.  
 DAUNT, J. C. 1951. *Proceedings of an International Conference on Low Temperature Physics* at Oxford, p. 157.  
 FLIPPEN, R. B. and FRIEDBERG, S. A. 1960. *J. Appl. Phys. Suppl.* **31**, 338.  
 GARRETT, C. G. B. 1951. *Proc. Roy. Soc. (London)*, A, **206**, 242.  
 GIJSMAN, H. M., POULIS, N. J., and VAN DEN HANDEL, J. 1959. *Physica*, **25**, 954.  
 GORTER, C. J. 1951. *Proceedings of an International Conference on Low Temperature Physics* at Oxford, p. 158.  
 GRACE, M. A., JOHNSON, C. E., KURTI, N., LEMMER, H. R., and ROBINSON, F. N. H. 1954. *Phil. Mag.* **45**, 1192.  
 GRIFFITHS, J. H. E. and OWEN, J. 1954. *Proc. Roy. Soc. (London)*, A, **226**, 96.  
 GRIFFITHS, J. H. E., OWEN, J., and WARD, I. M. 1953. *Proc. Roy. Soc. (London)*, A, **219**, 526.  
 GROTH, P. 1906. *Chemische Kristallographie. Bd 1* (Wilhelm Engelmann, Leipzig).  
 HOFFMANN, W. H. 1931. *Z. Krist.* **78**, 279.  
 MIEDEMA, A. R., VAN DEN BROEK, J., POSTMA, H., and HUISKAMP, W. J. 1959. *Physica*, **25**, 1177.  
 MIZUMO, J., UKEI, K., and SUGAWARA, T. 1959. *J. Phys. Soc. Japan*, **14**, 383.  
 MORIYA, T. 1956. *Progr. Theoret. Phys.* **16**, 641.  
 OHTSUBO, A., HASEDA, T., and KANDA, E. 1958. *Programme of the Kamerlingh Onnes Conference on Low Temperature Physics*, Leiden, p. 49.  
 PAULING, L. 1930. *Z. Krist.* **72**, 482.  
 POPPEMA, O. J. 1954. *Thesis Groningen*.  
 SHULMAN, R. G. and JACCARINO, V. 1958. *Phys. Rev.* **109**, 1084.  
 TUTTON, A. E. H. 1905. *J. Chem. Soc. Trans.* **87**, 1123.  
 ——— 1913. *Proc. Roy. Soc. (London)*, A, **88**, 361.  
 ——— 1916. *Phil. Trans. Roy. Soc. London*, A, **216**, 1.

# ELASTIC WAVES IN TRIGONAL CRYSTALS<sup>1</sup>

G. W. FARRELL<sup>2</sup>

## ABSTRACT

In non-isotropic single crystals the normals to the wavefronts of elastic waves are not colinear with the vectors representing either the energy flow or the particle displacement. Calculations have been carried out on the propagation characteristics of sound waves in two particular trigonal crystals,  $\alpha$ -quartz and sapphire.

The development of the eigenvalue equation for the velocity and the formulae for the components of the displacement and energy-flow vectors are summarized. The assumption that the wave has a plane wavefront normal to a given direction leads to three solutions, one representing a quasi-longitudinal wave and the other two representing quasi-transverse waves. The velocities of propagation, directions of displacement, and directions of energy flow for the three waves have been calculated for many orientations of the wave normal. Detailed results for propagation near one of the pure-mode axes are presented.

In the propagation of plane elastic waves in anisotropic but homogeneous solids, the phase velocity varies from direction to direction. In general the energy flow is not along the normal to the planes of constant phase nor is the displacement normal or parallel to these wavefronts. The outline of the calculation of the velocities and displacements for non-dispersive materials is given in the fascinating transcript of Lord Kelvin's Baltimore Lectures (1904) and the derivation of expressions for the direction of energy flow is given by Love (1927). These developments are summarized below in a more compact notation to define the various symbols used in the computations on specified trigonal crystals.

The interest in the two materials considered here arises from the experiments on the propagation of high-frequency sound waves in quartz by Bömmel and Dransfeld (1958, 1959) and Jacobsen (1959), and from the experiments on spin-lattice relaxation in ruby by Tucker (1960). Both of these materials,  $\alpha$ -quartz and ruby (sapphire or corundum with chromium impurities), form trigonal crystals. Examples of propagation in cubic and hexagonal crystals have been considered previously by Musgrave (1954a, b) and Miller and Musgrave (1956), the results being presented in the form of velocity, inverse, and wave surfaces.

## 1. GENERAL ANISOTROPIC CRYSTALS

For non-dispersive, homogeneous crystals of lowest symmetry the generalized Hooke's law relating the stress and the strain tensors may be written as

$$(1) \quad \sigma_{ij} = c_{ijkl}\epsilon_{kl} \quad (i, j, k = 1, 2, 3).$$

<sup>1</sup>Manuscript received July 21, 1960.

Contribution from the Eaton Electronics Research Laboratory, McGill University, Montreal. This research was supported by the Defence Research Board and the National Research Council.

<sup>2</sup>Temporary address: Clarendon Laboratory, Oxford, England.

A detailed summary of the repeated subscript notation for summation is given in Nye (1957). The components of the strain are related to the particle displacement  $\bar{u}$  by

$$(2) \quad \epsilon_{kl} = \frac{1}{2} \left( \frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right)$$

where  $\bar{u} = \bar{u}_1 u_1 + \bar{u}_2 u_2 + \bar{u}_3 u_3 = \bar{u}_i u_i$ , and  $\bar{u}_1, \bar{u}_2, \bar{u}_3$  are unit vectors along the Cartesian axes  $x_1, x_2, x_3$  respectively. The equations of motion satisfied by the components of the stress tensor and the particle displacements in a medium of density  $\rho$  are given by

$$(3) \quad \frac{\partial \sigma_{ij}}{\partial x_i} = \rho \ddot{u}_j.$$

Substituting equations (1) and (2) into (3) eliminates the stress components so that the equations of motion become

$$(4) \quad \frac{1}{2} c_{ijkl} \frac{\partial}{\partial x_i} \left( \frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right) = \rho \ddot{u}_i.$$

It is now assumed that the solution of these equations is a plane wave with constant amplitude  $\bar{A}$ , propagation factor  $k$ , and wavefronts normal to a vector with direction cosines  $l_1, l_2$ , and  $l_3$ ; that is

$$(5) \quad \bar{u} = \bar{A} \exp [ik(l_1 x_1 + l_2 x_2 + l_3 x_3 - vt)]$$

where  $\bar{A} = \bar{u}_i \alpha_i$ .

Here  $v$  is the phase velocity  $\omega/k$  in the direction given by  $l_i$ . When the assumed solution is substituted into equation (4) the following set of homogeneous equations in the components of the displacement result:

$$(6) \quad (\Gamma_{jk} - \delta_{jk} \rho v^2) \alpha_j = 0$$

with

$$\Gamma_{jk} = \frac{1}{2} l_i l_j (c_{ijkl} + c_{ijlk})$$

and

$$\delta_{jk} = \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases}$$

These equations have a non-trivial solution only if the secular equation

$$(7) \quad |\Gamma_{ij} - \delta_{ij} \rho v^2| = 0$$

is satisfied. Equation (7) is cubic in  $\rho v^2$  with three real roots so that for each corresponding value of  $v$  there is an allowed solution for which the relative values of the displacement components can be determined from equation (6). The displacement vectors are

$$(8) \quad \bar{A}_1 = C_1 \bar{u}_i a_i, \quad \bar{A}_2 = C_2 \bar{u}_i b_i, \quad \text{and} \quad \bar{A}_3 = C_3 \bar{u}_i c_i$$

associated with the three roots  $\rho v_1^2$ ,  $\rho v_2^2$ , and  $\rho v_3^2$  respectively. The  $C$ 's are constants dependent on the excitation and

$$(9) \quad \frac{a_2}{a_1} = \frac{\Gamma_{23}(\Gamma_{11}-\rho v_1^2) - \Gamma_{12}\Gamma_{13}}{\Gamma_{13}(\Gamma_{22}-\rho v_1^2) - \Gamma_{12}\Gamma_{23}}; \quad \frac{a_3}{a_1} = \frac{\Gamma_{23}(\Gamma_{11}-\rho v_1^2) - \Gamma_{12}\Gamma_{13}}{\Gamma_{12}(\Gamma_{33}-\rho v_1^2) - \Gamma_{23}\Gamma_{13}}$$

and

$$a_1^2 + a_2^2 + a_3^2 = 1$$

with similar expressions involving  $v_2$  and  $v_3$  for the  $b_i$  and the  $c_i$ .

The displacement vector corresponding to the largest root  $\rho v_1^2$  is more or less in the direction of the normal to the wavefront and thus represents a quasi-longitudinal or compressional wave, whereas the other two vectors of the mutually orthogonal triad are quasi-transverse or shear waves. In the following sections these waves will be referred to simply as longitudinal and transverse.

The vector whose components parallel to the axes are given by

$$(10) \quad E_i = \dot{u}_i \sigma_{ij}$$

represents the energy flow per unit time across a surface of unit area normal to this vector (Love, Chapter 7). When equations (1) and (2) and (8) are substituted into (10), the result averaged over a cycle gives an analogue to Poynting's vector for the direction of energy flow. The components of this vector are

$$(11) \quad S_i = \frac{1}{2} C_1^2 \frac{\omega^2}{v_1} c_{ijk} a_j (l_i a_k + l_k a_i)$$

with similar expressions for the other two waves.

## 2. TRIGONAL CRYSTALS

When the crystal symmetry is taken into account the number of stiffness constants is greatly reduced from the 81 indicated in equation (1). The two trigonal crystals of interest here,  $\alpha$ -quartz and sapphire, each have a threefold axis of symmetry which is taken as the  $x_3$  or  $Z$ -axis and three twofold axes at right angles to the  $Z$ -axis. One of the diad axes is taken as the  $x_1$  or  $X$ -axis. The co-ordinate system is illustrated in Fig. 1.

For crystal of this class, the  $\Gamma$ 's of equation (6) reduce to

$$(12) \quad \begin{aligned} \Gamma_{11} &= l_1^2 c_{11} + \frac{l_2^2}{2} c_{11} - \frac{l_2^2}{2} c_{12} + l_3^2 c_{44} + 2l_2 l_3 c_{14} \\ \Gamma_{22} &= \frac{l_1^2}{2} c_{11} - \frac{l_1^2}{2} c_{12} + l_2^2 c_{11} + l_3^2 c_{44} - 2l_2 l_3 c_{14} \\ \Gamma_{33} &= (l_1^2 + l_2^2) c_{44} + l_3^2 c_{33} \\ \Gamma_{12} &= 2l_1 l_3 c_{14} + \frac{1}{2} l_1 l_2 (c_{11} + c_{12}) \\ \Gamma_{13} &= l_1 l_3 (c_{44} + c_{13}) + 2l_1 l_2 c_{14} \\ \Gamma_{23} &= (l_1^2 - l_2^2) c_{14} + l_2 l_3 (c_{13} + c_{44}) \end{aligned}$$

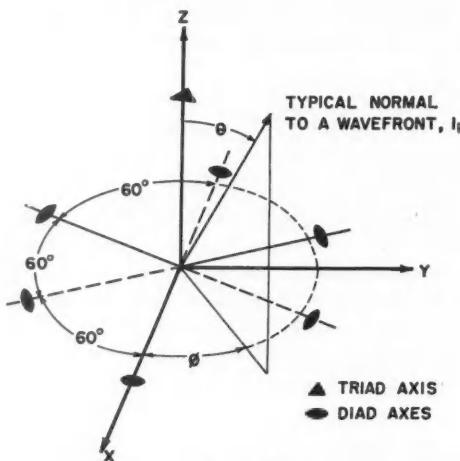


FIG. 1. Cartesian co-ordinate system for the trigonal crystals showing equivalent axes in the  $NY$  plane.

where the stiffness tensor  $c_{ijkl}$  has been expressed in the more common matrix form  $c_{mn}$  (see, for example, Nye, Chapter 8).

The values used for the stiffness constants in the subsequent calculations are shown in Table I\* (Bhimasenachar 1949, 1950; Cady 1946), and the effects of recently published values for sapphire (Wachtman *et al.* 1960) are discussed in the Appendix.

TABLE I  
Stiffness constants

Crystal	$\rho$ ( $\text{kg}/\text{m}^3$ )	$c_{11}$	$c_{22}$	$c_{44}$ (in units of $10^{11}$ newtons/ $\text{m}^2$ )	$c_{13}$	$c_{12}$	$c_{14}$
Quartz	2650	0.8694	1.0680	0.5762	0.0696	0.1560	0.1743
Sapphire	4000	4.65	5.63	2.33	1.24	1.17	1.01

For a given direction of the normal to the wavefront expressed by the direction cosines  $l_i$ , the  $\Gamma$  coefficients can be evaluated and hence the allowed velocities determined from the roots of the cubic equation (7). Various cross sections of the velocity surfaces for these two materials are shown on Fig. 2. Similar curves for quartz are given by Koga (1936) and Bechmann (1935) and are reproduced in Cady (1946). Because of the crystal symmetry, the  $XZ$  and  $YZ$  cross sections are repeated each  $120^\circ$ , thus these two cross sections

\*The choice of axes for right-handed  $\alpha$ -quartz agrees with that in Cady (1946). Using the axes defined in the "I.R.E. standards on piezoelectric crystals" (1949) reverses the sign of  $c_{14}$  (Koga, Aruga, and Yoshinaka 1958).

FIG. 2. Cross sections of the velocity surface for quartz (above) and sapphire (below). The small circles are values calculated using Wachtman's constants, see Appendix. The dashed curves are calculated from the perturbation method of Waterman.

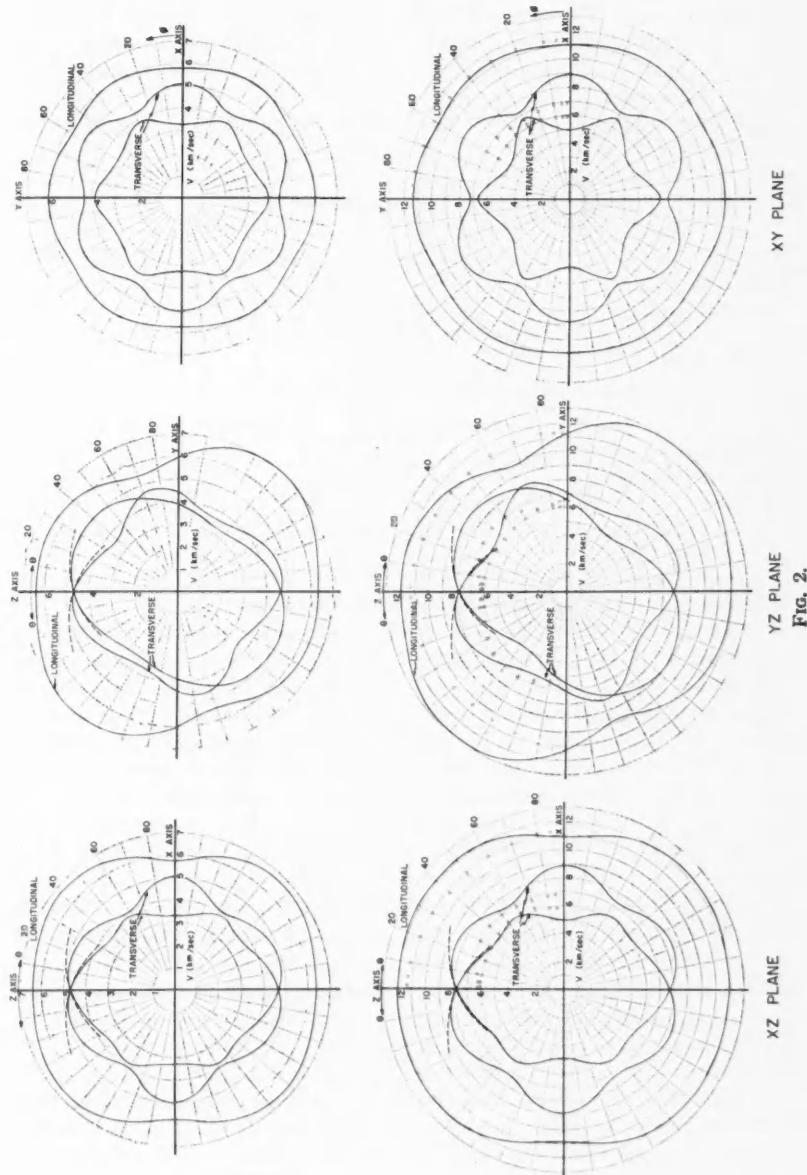


FIG. 2.

are sufficient to determine the shape of the velocity surfaces at 30° increments in azimuth angle. The points at which the velocities of the two transverse waves are equal are discussed below. The dotted curves shown in Fig. 2 are calculated from formulae derived by Waterman (1959) for propagation close to pure-mode axes.

For these trigonal crystals the components of Poynting's vector are of the form

$$(13) \quad \begin{aligned} S_x &= C_1^2 \frac{\omega^2}{v_1} (\beta_1 a_x + \beta_6 a_y + \beta_5 a_z) \\ S_y &= C_1^2 \frac{\omega^2}{v_1} (\beta_6 a_x + \beta_2 a_y + \beta_4 a_z) \\ S_z &= C_1^2 \frac{\omega^2}{v_1} (\beta_5 a_x + \beta_4 a_y + \beta_3 a_z) \end{aligned}$$

where

$$\begin{aligned} \beta_1 &= c_{11}l_1 a_x + c_{12}l_2 a_y + c_{13}l_3 a_z + c_{14}(l_3 a_y + l_2 a_z) \\ \beta_2 &= c_{12}l_1 a_x + c_{11}l_2 a_y + c_{13}l_3 a_z - c_{14}(l_3 a_y + l_2 a_z) \\ \beta_3 &= c_{13}l_1 a_x + c_{12}l_2 a_y + c_{33}l_3 a_z \\ \beta_4 &= c_{14}l_1 a_x - c_{14}l_2 a_y + c_{44}(l_3 a_y + l_2 a_z) \\ \beta_5 &= c_{44}(l_1 a_x + l_3 a_z) + c_{14}(l_2 a_x + l_1 a_y) \\ \beta_6 &= c_{14}(l_1 a_x + l_3 a_z) + \frac{1}{2}(c_{11} - c_{12})(l_2 a_x + l_1 a_y) \end{aligned}$$

again with corresponding expressions for the other two waves.

The data for the displacement and energy flow were calculated as the points of intersection of the corresponding vectors with a unit sphere for each direction of the wave normal. These three normalized vectors are designated as  $a_i$ ,  $S_i$ , and  $l_i$  respectively. For presentation here the unit sphere is mapped by a central projection (Steers 1942) onto a right hexagonal prism as shown in Fig. 3. With this form of projection the crystal symmetry is preserved and all great circles project as straight lines and conversely all straight lines represent segments of great circles. Thus the line joining the point representing a given  $l_i$  with the point for the corresponding  $S_i$  (or  $a_i$ ) is a projected segment of the great circle containing these two vectors. The typical vector shown on Fig. 3 is  $l_i = \sin 50^\circ, 0^\circ, \cos 50^\circ$ . If  $\Delta\theta$  is the difference in zenith angle  $\theta$ , and  $\Delta\phi$  the difference in azimuth angle  $\phi$  between two vectors, then the contained angle  $\gamma$  is given by

$$(14) \quad \cos \gamma = \cos \theta_2 \cos \theta_1 + \sin \theta_2 \sin \theta_1 \cos \Delta\phi$$

or if  $\gamma$  is small

$$(15) \quad \gamma \cong [(\Delta\theta)^2 + (\Delta\phi \sin \theta)^2]^{\frac{1}{2}}$$

Figures 4a and 4b show the directions of displacement in the longitudinal wave for different directions of the normal to the wavefront. On this scale the energy-flow vectors, or rays, for longitudinal waves are almost coincident with the vectors for displacement and have not been shown separately. This drawing and the subsequent Figs. 6, 7, and 8 are to be read as follows: for a

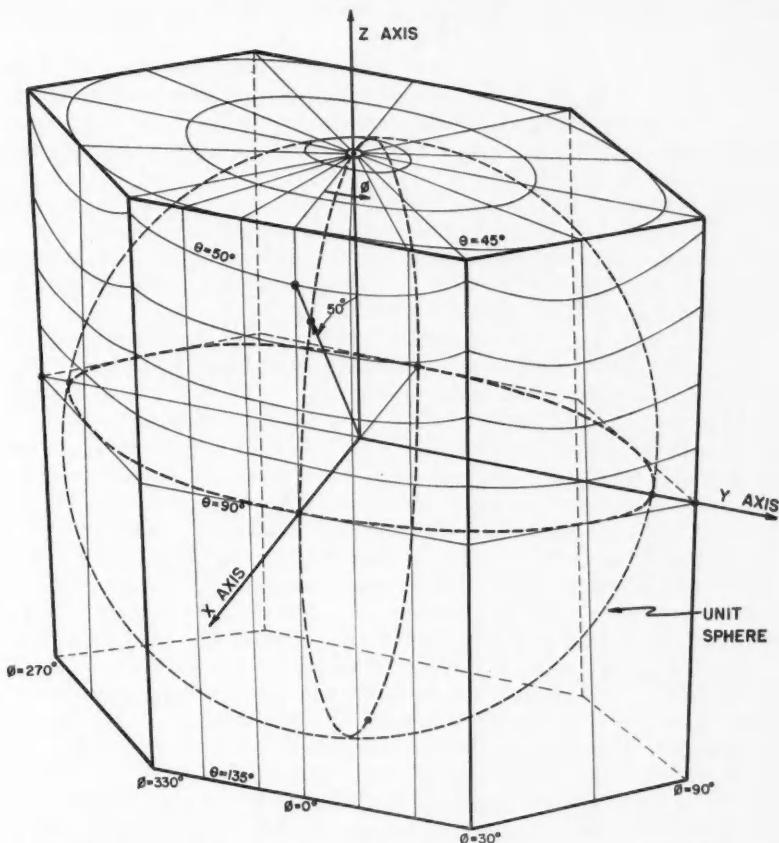


FIG. 3. Projection of the unit circle onto a right hexagonal prism. This prism is then unfolded for the presentation of the data in Figs. 4, 6, etc.

selected  $l_i$  as indicated by one of the small circles, the corresponding  $S_i$  (or  $a_i$ ) is given by the cross which is connected to this circle by a straight line; the latter line also represents the great circle containing the  $l_i$  and  $S_i$ . To aid in interpolation between the plotted values, the loci of the rays for wave normals in specified planes are shown by broken lines. The  $X$  and  $Z$  axes are pure-mode axes for longitudinal waves, in that for these directions the  $l_i$ ,  $a_i$ , and  $S_i$  are colinear, while the  $YZ$  and equivalent planes are planes of mirror symmetry. The maximum angle between an  $l_i$  and the corresponding  $S_i$  is about  $25^\circ$ .

The longitudinal displacement and energy-flow vectors for wave normals in the  $YZ$  plane lie in the  $YZ$  plane itself and their directions are shown by one of the curves in each of Figs. 5a and 5b. In this illustration, positive values of zenith angle are used for vectors with positive  $y$ -components. Other pure-mode

axes for longitudinal wave propagation (Borgnis\* 1955) are seen on these diagrams and are marked also on Fig. 4.

The small pairs of arrows shown at some of the crosses of Fig. 4 give the directions of the displacements for the two corresponding transverse waves. The arrows indicate the great circle directions from the tip of the longitudinal displacement vector to the tips of the mutually perpendicular displacement vectors of the transverse waves.

While only about an octant of the unit sphere is portrayed in Fig. 4, symmetry conditions permit the extension of these results, in conjunction with Fig. 5, to give rays and displacements for wave normals in the full equatorial plane and in meridional planes at  $30^\circ$  increments in azimuth. Directions equivalent to the various axes are marked on Fig. 4.

Figures 5 and 6 show the energy-flow directions for the transverse waves using the same representation as for the longitudinal case. Here it is seen that in some regions the  $S_i$  depart from the corresponding  $l_i$  by angles as great as  $50^\circ$ . The  $X$ -axis is a pure-mode axis for transverse waves also. The  $Z$ -axis is a degenerate axis for transverse waves in that the two velocities are equal and for  $l_i$  coincident with this axis one gets the conical energy flow discussed by Miller and Musgrave (1956) and further interpreted by Waterman (1959). The interesting behavior of the rays for these transverse waves as the wave normal departs from the  $Z$ -axis in various planes is shown in Figs. 5 and 6. Similar behavior is evident in the vicinity of the direction of equal velocities in the  $YZ$  plane (see Fig. 2). For wave normals in the latter plane one of the transverse waves has its displacement vector perpendicular to the  $YZ$  plane while the other has its displacement in the plane. Within the interval marked on Fig. 5 the wave with the larger velocity has the perpendicular displacement, whereas outside of this range the wave with the smaller velocity has the perpendicular displacement. For some directions considerably removed from those of equal velocity, for example, near  $l_i$  given by  $\theta = 55^\circ$ ,  $\phi = 0^\circ$  and  $\theta = 90^\circ$ ,  $\phi = 30^\circ$ , the energy-flow vector varies widely for small changes in  $l_i$ . The latter directions correspond to promontories shown on the smaller velocity curves in the  $XZ$  and  $XY$  planes respectively of Fig. 2.

The other pure-mode axes, for which the rays and the wave normals are colinear and the displacement is normal to the common direction, are shown on Figs. 5 and 6. These directions at  $\theta = -59^\circ$  and  $\theta = +31^\circ$  in quartz are termed the  $AC$  and  $BC$  axes respectively.

For wave normals almost colinear with the  $X$ -axis, the propagation characteristics of the longitudinal wave are shown in Fig. 7. Here the directions of energy flow are given for wave normals lying in cones of semi-angle  $2^\circ$  and  $1^\circ$ . The displacements for the  $2^\circ$  case are shown by the small triangles. It is seen that in experiments on sound propagation down  $X$ -cut rods, such as those of Bömmel and Dransfeld (1958) or Jacobsen (1959), a misalignment of  $1^\circ$  between the rod axis and the  $X$ -axis can cause an angle as large as  $6^\circ$  or as small as  $0.5^\circ$  between the ray and the rod axis depending on the direction of the misorientation.

\*Borgnis defined his pure-mode axes for longitudinal waves by requiring only that  $l_i$  and  $a_i$  be colinear. For the cases considered here, when  $l_i$  is colinear with  $a_i$  for longitudinal waves,  $S_i$  is also colinear with  $l_i$ .

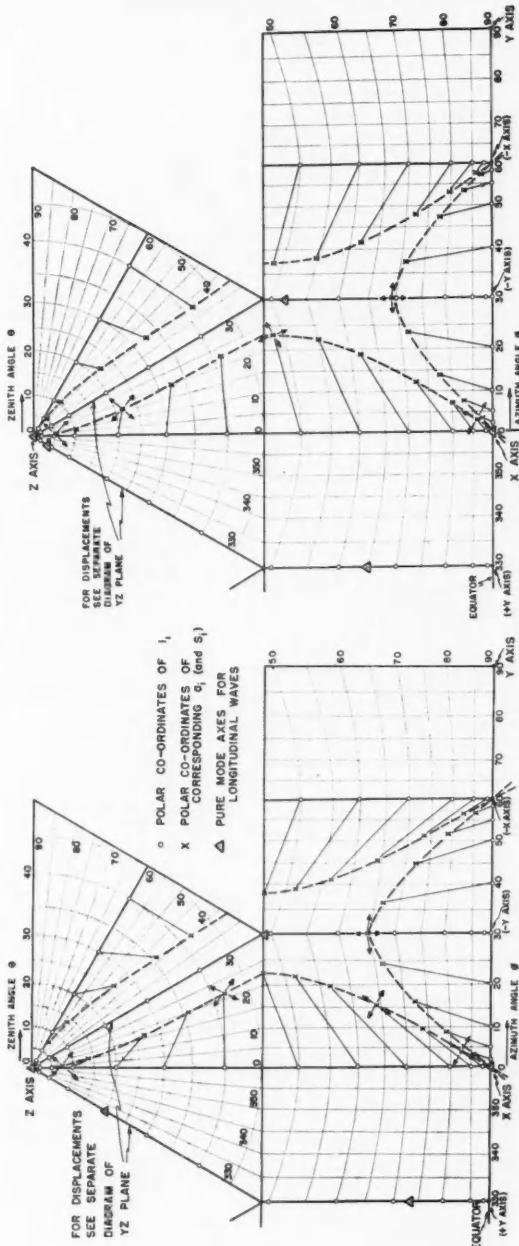


FIG. 4. Displacement vectors for the longitudinal wave in quartz (a) and sapphire (b). Energy-flow vectors are essentially colinear with the displacement vectors.

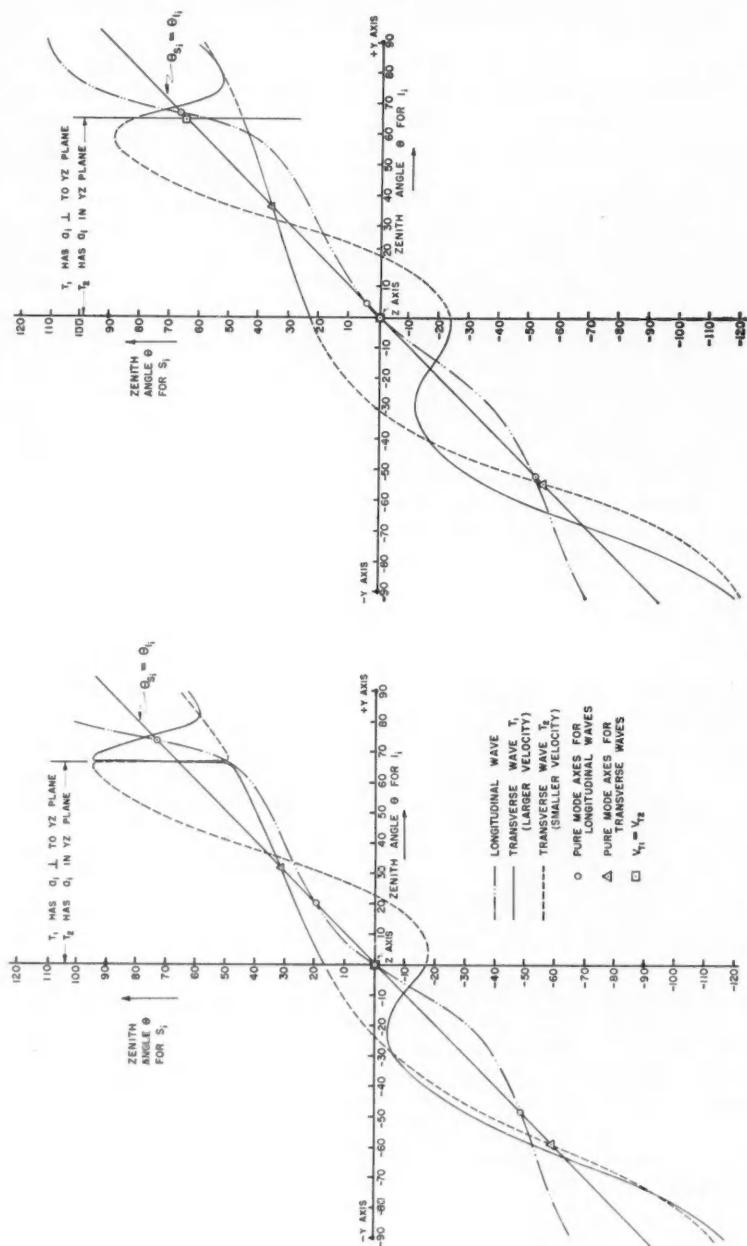


FIG. 5.

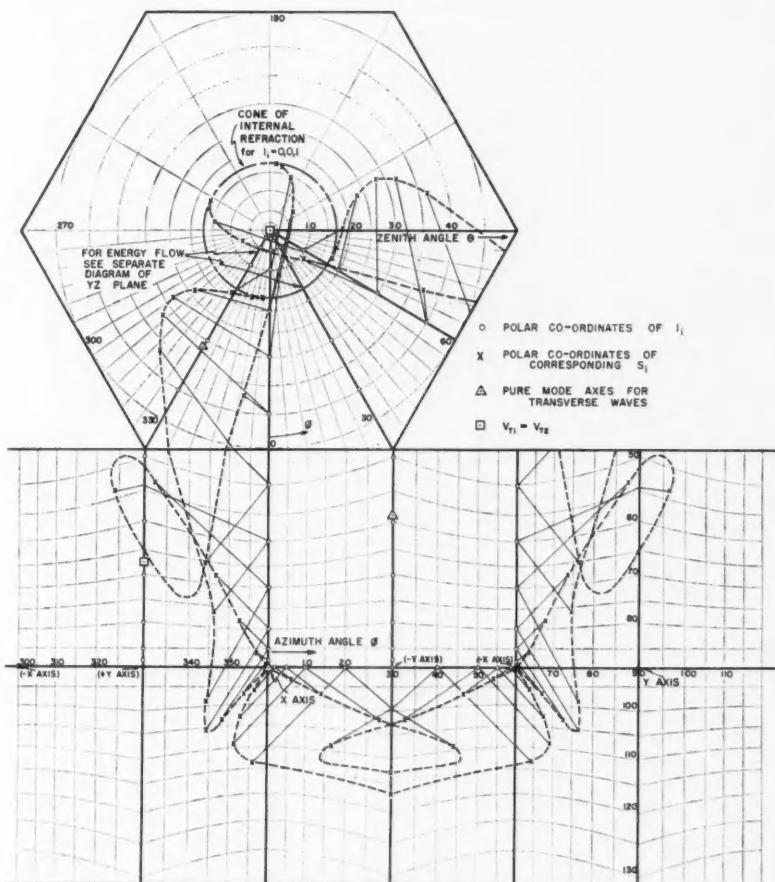


FIG. 6a. Energy-flow vectors for the transverse waves in quartz.

As noted previously, the  $X$ -axis is also a pure-mode axis for transverse waves and the displacement directions for wave normals along this axis are shown by the arrows at the origins in Fig. 8.  $T_1$  represents the transverse wave with the larger velocity. The loci of the energy-flow directions for wave normals lying in a  $2^\circ$  cone about the  $X$ -axis are shown by the broken lines. Here the angle

FIG. 5. Energy-flow directions for wave normals in the  $YZ$  plane for quartz (left) and sapphire (right). The abscissa is the zenith angle of a wave normal and the ordinate is the zenith angle of the corresponding energy-flow vector. For points on the  $45^\circ$  line the two vectors are colinear.

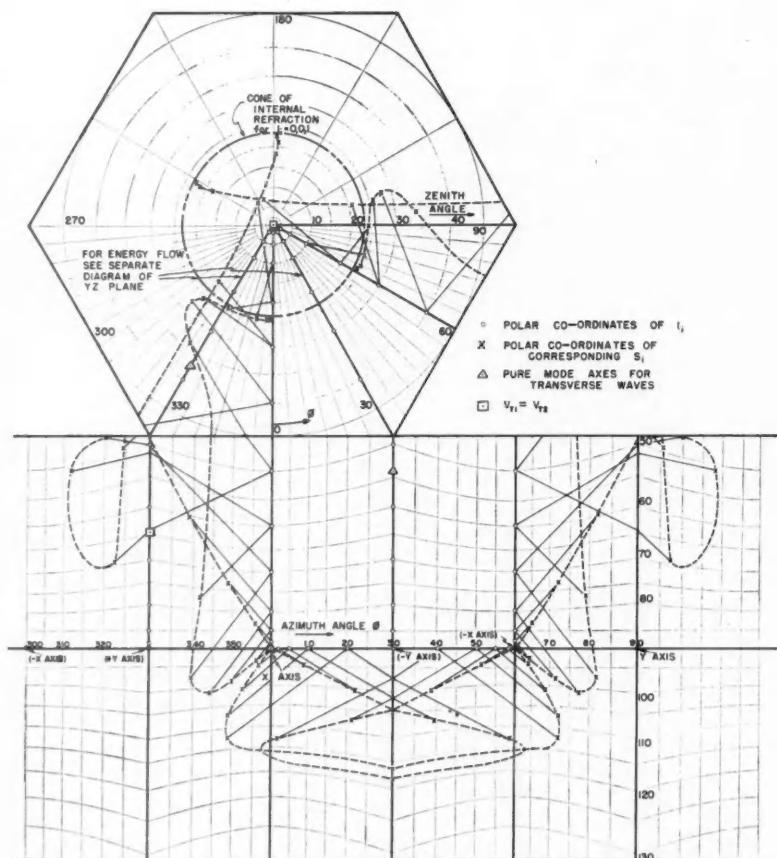


FIG. 6b. Energy-flow vectors for the transverse waves in sapphire.

between the wave normal and the corresponding ray varies from about  $12^\circ$  to less than  $1^\circ$  depending on the wave considered and the location of the  $l_i$  on the cone. The points on the axes representing  $l_i$  in the  $XY$  and  $XZ$  planes are the same as the corresponding points in the smaller scale drawings of Fig. 6.

FIG. 7. Energy-flow and some displacement directions for the longitudinal wave in quartz (left) and sapphire (right) for wave normals  $1^\circ$  and  $2^\circ$  from the  $X$ -axis.

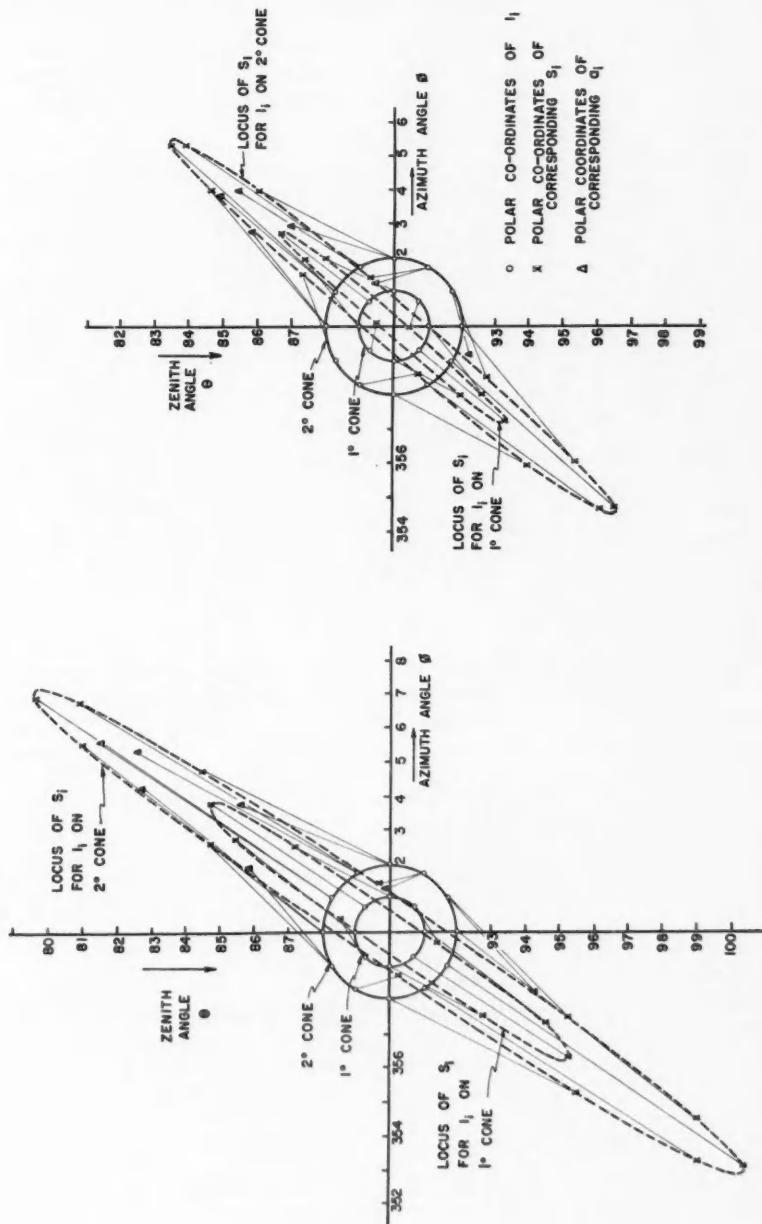


Fig. 7.

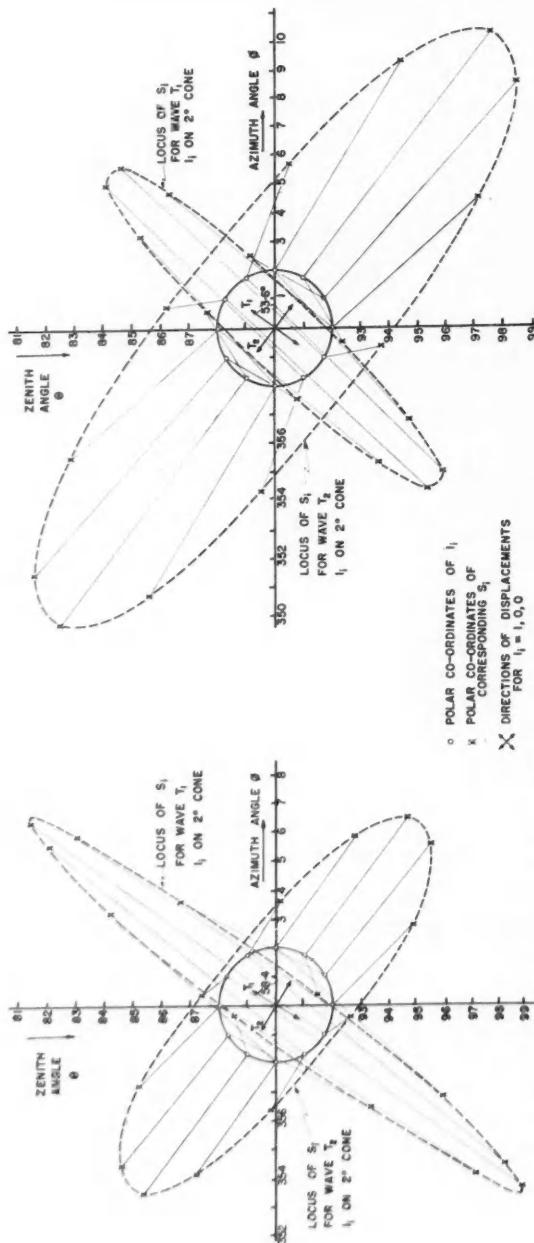


Fig. 8. Energy-flow vectors for the two transverse waves in quartz (left) and sapphire (right) for wave normals  $2^\circ$  from the  $X$ -axis.

## ACKNOWLEDGMENTS

The author appreciates the encouragement offered by Prof. G. A. Woonton and the services provided by the personnel of the McGill Computation Center in assisting with the programming of these calculations.

## REFERENCES

BECHMANN, R. 1935. *Z. tech. Physik*, **16**, 525.  
 BHIMASENACHAR, J. 1949. *Current Sci. (India)*, **18**, 372.  
 ——— 1950. *Proc. Natl. Inst. Sci. India*, **16**, 241.  
 BORGNISS, F. E. 1955. *Phys. Rev.* **98**, 1000.  
 BÖMMEL, H. E. and DRANSFIELD, K. 1958. *Phys. Rev. Letters*, **1**, 234.  
 ——— 1959. *Phys. Rev. Letters*, **2**, 298.  
 CADY, W. G. 1946. *Piezoelectricity* (McGraw Hill, New York), Chaps. V and VI.  
 JACOBSEN, E. H. 1959. *Phys. Rev. Letters*, **2**, 249.  
 KELVIN (Lord). 1904. *Baltimore lectures* (Cambridge University Press).  
 KOGA, I. 1936. *Proc. I.R.E.*, **24**, 510.  
 KOGA, I., ARUGA, M., and YOSHINAKA, Y. 1958. *Phys. Rev.* **109**, 1467.  
 LOVE, A. E. H. 1927. *On the mathematical theory of elasticity* (Cambridge University Press; also Dover Publications, New York 1944).  
 MILLER, G. F. and MUSGRAVE, M. J. P. 1956. *Proc. Roy. Soc. (London)*, A, **236**, 352.  
 MUSGRAVE, M. J. P. 1954a. *Proc. Roy. Soc. (London)*, A, **226**, 339.  
 ——— 1954b. *Proc. Roy. Soc. (London)*, A, **226**, 356.  
 NYE, J. F. 1957. *Physical properties of crystals* (Oxford Clarendon Press), Chaps. I and VIII.  
 STANDARDS ON PIEZOELECTRIC CRYSTALS. 1949. *Proc. I.R.E.* **37**, 1378.  
 STEERS, J. A. 1942. *Study of map projections* (University of London Press), Chap. I.  
 TUCKER, E. B. 1960. *Detroit Mtg. Am. Phys. Soc. Paper F4*, March 1960.  
 WACHTMAN, J. B., JR., TEFFT, W. E., LAM, D. G., and STINCHFIELD, R. P. 1960. *J. Research Natl. Bur. Standards*, **64**, A, 213.  
 WATERMAN, P. C. 1959. *Phys. Rev.* **113**, 1240.

## APPENDIX

Since the preparation of the original drawings for this paper Wachtman *et al.* (1960) have published results on the measurements of the elastic constants of artificial sapphire. These values which are listed in Table II differ appreciably from those of Bhimasenachar for natural sapphire crystals as given in Table I.

TABLE II  
 Stiffness constants for sapphire according to Wachtman *et al.* (1960) (in  $10^{11}$  newtons/m<sup>2</sup>)

$c_{11}$	$c_{33}$	$c_{44}$	$c_{12}$	$c_{13}$	$c_{14}$
4.968	4.981	1.474	1.636	1.109	-0.235

In particular the second set of values indicates a higher degree of isotropy than the earlier values. This can be seen as follows: in trigonal crystals of class  $\bar{3}m$ , the stiffness constants listed in the tables are independent, and the remaining non-zero elements of the symmetrical matrix satisfy

$$(A.1) \quad \begin{aligned} c_{22} &= c_{11}; & c_{23} &= c_{13}; & c_{55} &= c_{44} \\ -c_{24} &= c_{56} = c_{14}; & \text{and} & & c_{66} &= \frac{1}{2}(c_{11} - c_{12}) \end{aligned}$$

which are the relations used in deriving equation (12) from equation (6). The additional relations which must be more closely satisfied as a trigonal crystal approaches isotropy are

$$(A.2) \quad \begin{aligned} c_{33} &= c_{11}; & c_{44} &= \frac{1}{2}(c_{11} - c_{12}); \\ c_{13} &= c_{12} & \text{and} & \quad c_{14} = 0 \end{aligned}$$

Comparison of the two sets of values for the  $c$ 's shows that those of Table II come much closer to satisfying equations (A.2) than those of Table I.

The more isotropic nature of the propagation using Wachtman's values is borne out by detailed calculations. For example, the small circles shown on Fig. 2 indicate points on the velocity surfaces obtained with these values, whereas Fig. A1 is a recalculated version of Figs. 4b and 6b showing the

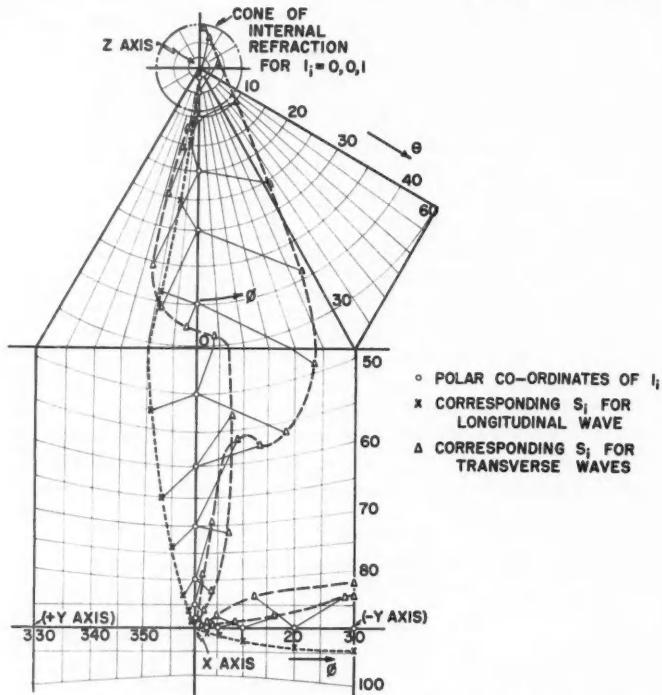


FIG. A1. Energy-flow vectors for longitudinal and transverse waves in sapphire using Wachtman's constants. The drawing has mirror symmetry about the plane  $\phi = 30^\circ$  which is equivalent to a  $YZ$  plane.

directions of energy flow. The latter values give degenerate axes for the transverse waves near  $\theta = -65^\circ, -35^\circ$ , and  $75^\circ$  in the  $YZ$  plane as well as the  $Z$ -axis itself. Pure-mode axes occur in the  $YZ$  plane at  $\theta = -50^\circ$ , and  $37^\circ$  for the longitudinal wave and at  $\theta = -51^\circ, -8^\circ$ , and  $38^\circ$  for the transverse waves. Comparison of Figs. 4b and 6b with Fig. A1 shows smaller angles on the latter drawing between the energy-flow vectors and the corresponding wave normals; moreover, in general the energy-flow vectors are in radically different directions from those in the former case.

## SCATTERING BY A GRATING. I<sup>1</sup>

R. F. MILLAR

### ABSTRACT

Consideration is given to scattering of plane waves by a transmission grating upon whose elements the wave function vanishes. By an application of Green's theorem, the problem is formulated in terms of integral equations for a finite, or infinite, number of elements. The equations are put in a form more suitable for the study of interaction phenomena by the subtraction of a certain series. The modified equations then correspond, more or less, to the excitation of each element of the grating by the incident field, and two plane waves propagated in opposite directions along the grating. A solution is here attempted only for the infinite grating of identical elements. Attention is confined to the region of "Rayleigh wavelengths", where interaction is important, and the variation in spectral intensity for a grating of elements of arbitrary form is discussed in a semiquantitative manner; the existence of anomalies is inferred. An explicit solution is obtained for small scatterers of elliptical cross section, and the behavior of the spectral intensity is considered in some detail.

### 1. INTRODUCTION

The scattering of plane waves by gratings has been considered frequently in the scientific literature. Until the turn of the century, the notions of Fresnel and Huygens had appeared adequate to predict the observed scattering properties of optical reflection gratings, but it was realized by Rayleigh that the Fresnel theory could be justified only if the grating interval was large in comparison with the wavelength of the incident radiation. For smaller separation, a more rigorous approach was indicated, and Rayleigh (1907) provided a perturbation-type analysis for a perfectly conducting, infinite reflection grating, under the hypothesis that the depth of the corrugations was small in comparison with the wavelength.

For the case in which the magnetic vector of the incident wave was parallel to the rulings of the grating (which here will be referred to as the case of *H*-polarization) it was observed that Rayleigh's method of approximation failed when any of the scattered plane waves was propagated tangentially along the surface, the inference drawn being the probable existence of abnormalities in the spectra. No such behavior was apparent in the solution for *E*-polarization.

About this time, abrupt and unexpected changes (or anomalies) in the intensity of the spectra of an optical reflection grating had been reported by Wood (1902). The "Wood anomalies" were then observed only when the incident light contained an *H*-polarized component, and at precisely the "Rayleigh wavelengths" for which Rayleigh's approximation failed. Later, Wood (1912, 1935) extended his observations, and noted the presence of relatively weak anomalies in the case of an *E*-polarized incident field. More

<sup>1</sup>Manuscript received September 6, 1960.

Contribution from the Radio and Electrical Engineering Division, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 6092.

recently, experimental evidence for the existence of anomalies in the latter case has been produced by Palmer (1952), who verified that anomalous behavior occurred at wavelengths for which the Rayleigh approximation failed. Palmer concluded that the theoretical non-existence of anomalies for *E*-polarization was a consequence of the assumption (in the Rayleigh analysis) of shallow corrugations. Meecham (1956) presented a variational treatment of the problem for a perfectly conducting reflection grating which indicated weak anomalies in the case of *E*-polarization. It will be seen later that for any transmission grating with a sufficiently large number of elements on which are satisfied the assumed boundary values, anomalous behavior is to be expected for either polarization; for a reflection grating anomalous effects, although present, depend more strongly on polarization.

In recent years, the stimulus of microwave technology has led to the appearance of many more theoretical treatments of grating problems.\* Infinite transmission gratings of thin wires have received frequent attention (see, for example, Wait 1955), in general subject to the assumptions of a line current along the axis of each wire, and separation small in comparison with wavelength; hence the possibility of anomalous behavior does not arise. (Of course, the infinite grating is a mathematical idealization, and in microwave practice the number of elements is considerably smaller than in optical applications. Thus interaction effects are less pronounced, and anomalous behavior is not so likely to be observed at microwave frequencies.)

The first successful attempt to explain the Wood anomalies appears to have been made by Artmann (1942), who considered the scattering of an *H*-polarized plane wave by an infinite reflection grating. In the analysis it was assumed that the periodic function which defined the surface was a small perturbation on a plane. Artmann obtained explicit expressions for the complex amplitudes of the various spectral plane waves in the neighborhood of an anomaly, in terms of the geometry of the reflecting surface. (The validity, in the region between elements of the grating, of the discrete plane wave representation for the scattered field employed by Rayleigh (1907) and Artmann (1942) has been questioned by Lippmann (1953).)

More recently, the anomalous behavior of transmission and reflection gratings has been studied by Twersky (1952, 1956, 1957, 1958), and by Karp and Radlow (1956). Twersky (1952) applied a theory of multiple scattering to the finite transmission and reflection gratings, and in the later papers derived a "sum-integral" equation which related the amplitude scattered by a cylinder in the grating to the corresponding quantity for an isolated cylinder.

Because the anomalies are associated with waves propagated tangentially along the grating, it would seem logical, in an attempt to explain the phenomena, to incorporate this property in the initial formulation of the problem. This was done by Karp (1955), who showed that the field scattered by each cylinder was approximately that scattered by an isolated cylinder when excited by the incident and the tangentially propagating waves.

\*A survey of recent analytical work on multiple scattering has been conducted by Twersky (1960) in a report prepared for U.S. Commission VI, International Scientific Radio Union (URSI).

The analysis which follows is essentially an extension of some earlier work (Millar 1960, and henceforth referred to by (†)), in which was considered the scattering of a scalar plane wave by a row of  $N$  cylinders, subject to the boundary condition that either (i) the total wave function ( $u$ ) or (ii) its normal derivative ( $\partial u / \partial n$ ) vanished on the scatterers. In each case, a set of simultaneous, linear integral equations was derived for the unknown field distribution on the cylinders, and an approximate "power series" (perturbation-type) solution was obtained for elliptical scatterers much less than the wavelength in cross-sectional dimensions.

Although the rigorous solution of the integral equations (were such possible to obtain) would include all interaction and anomalous effects, it was found that the approximate, small-cylinder solution was valid only when interaction was not of primary importance. This restriction was embodied in the condition that  $(N-1)/(ka)$  be sufficiently small for the given solution to be considered a legitimate approximation. (Here  $a$  is the spacing between elements of the grating, and  $k$  ( $= 2\pi/\lambda$ ) the wave number.) The infinite grating was therefore by hypothesis excluded from consideration.

(In this connection, it may be worth while to note that a calculation of the first term in the above perturbation-type solution for the infinite grating would indicate nothing amiss. Nevertheless, the coefficients of higher-order terms become infinite at the Rayleigh wavelengths. The first term in the solution for problem (i) vanishes at the Rayleigh wavelengths, in agreement with the results of Wait (1955) and others. However, the divergent behavior of the higher-order terms leads one to treat the first with caution. It will be shown later by a more careful analysis that the first term in the appropriate solution does, indeed, vanish in these circumstances, while higher-order terms remain finite.)

It is the purpose of the present paper to demonstrate a method by which the effects of interaction are introduced explicitly into the integral equations for the two boundary-value problems mentioned above. In many respects, the notions involved are similar to the procedure adopted by Karp (1955) and Karp and Radlow (1956). However, it may be worth while to mention that no approximation is involved in obtaining the modified integral equations from the old; the new equations are rigorous, and possess an approximate solution for small scatterers more suited to the discussion of interaction effects than the original set. The number of cylinders may be finite or infinite—basically the same integral equations are found in either case, although the infinite grating is more amenable to analysis.

The paper is divided into two parts. The remainder of Part I consists of a fairly detailed analysis of the problem in which the wave function vanishes on the grating elements. In Section 2, the modified integral equations are derived, and formal solutions are given. The infinite (transmission) grating of identical cylinders is considered in Section 3, in which case periodicity of the structure reduces the number of unknowns to unity. A few observations of a semiquantitative nature are made in connection with gratings with elements of arbitrary shape. An approximate solution is obtained under the assumption of small, elliptic scatterers.

## 2. DERIVATION AND FORMAL SOLUTION OF THE INTEGRAL EQUATIONS ( $u = 0$ on the cylinders)

Let the cylinders be arranged in such a manner that their generators lie parallel to the  $x$ -axis of a rectangular Cartesian co-ordinate system  $(x, y, z)$  (see Fig. 1). Because of the two-dimensional nature of the problem, the  $x$

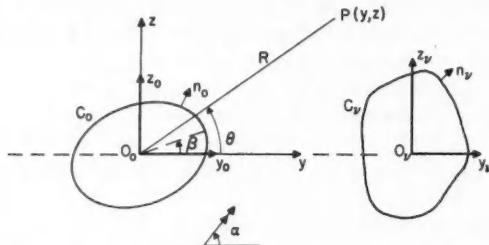


FIG. 1. Configuration of cylinders and symbols employed in the analysis.

co-ordinate will be ignored subsequently. For the present, it is not essential to specify the shape of any particular cylinder, no two of which need necessarily be alike, but it is assumed that some point interior to each (the center) lies on the positive  $y$ -axis, the origin of the co-ordinate system being located at the center of the zeroth cylinder  $C_0$ . The center  $O_v$  of the  $v$ th cylinder  $C_v$  has co-ordinates  $(va, 0)$  in the  $(y, z)$  system. Here  $a$  is the distance between centers of successive scatterers. In addition to the  $(y, z)$  co-ordinates, a  $(y_v, z_v)$  system (with origin at  $O_v$ ) is defined to specify points on each cylinder.

If the total wave function is  $ue^{-i\omega t}$ , where  $u = u^i + u^s$ , where  $u^s$  is the scattered field which satisfies an appropriate radiation condition, and where  $u^i = e^{ik(y \cos \alpha + z \sin \alpha)}$  is the incident plane wave whose propagation vector makes an angle  $\alpha$  with the positive  $y$ -axis, then an application of Green's theorem leads to an expression for the field at a point  $P(y, z)$  exterior to all the cylinders in terms of the boundary values of  $u$ ,  $\partial u / \partial n$  on the cylinders:

$$(1) \quad u(P) = u^i(P) + \frac{1}{4}i \sum \int_{C_v} \left[ u \frac{\partial}{\partial n_v} H(kr_v) - H(kr_v) \frac{\partial u}{\partial n_v} \right] ds_v.$$

Here  $n_v$  is the normal, outward drawn from  $C_v$ ,  $ds_v$  is the element of length along  $C_v$ ,  $H(kr_v)$  represents the Hankel function of the first kind of order zero, and

$$(2) \quad r_v^2 = [y - (va + y_v)]^2 + [z - z_v]^2.$$

The summation is performed over all  $v$ , the number of cylinders being finite, or infinite.

Because  $u = 0$  on the cylinders in the problem under consideration, equation (1) for  $u(P)$  becomes

$$(3) \quad u(P) = u^i(P) - \frac{1}{4}i \sum \int_{C_v} H(kr_v) \chi_v ds_v,$$

where  $\chi_r \equiv \partial u / \partial n_r$ . If  $P$  approaches a point  $P_\lambda$  on  $C_\lambda$ , then a set of integral equations is obtained for  $\chi_r$ :

$$(4) \quad \sum \int_{C_r} H(kr_{r\lambda}) \chi_r ds_r = -4iu^1(P_\lambda).$$

As in (†), lower case Greek subscripts will refer to points of integration, while upper case subscripts imply points fixed on the corresponding cylinders.

The set of equations (3†) analogous to equations (4), but involving the normal derivative of the scattered field on the cylinders, was solved approximately for a finite number ( $N$ ) of small, similarly oriented, elliptic scatterers. An examination of this solution leads one to conclude that when  $N$  is very large, the approximation may become invalid. This in turn may be traced to the behavior of

$$\sum H(kr_{r\lambda}) \chi_r.$$

For, if  $N$  is large, the function  $\chi_r$  differs from  $\chi_{r+\mu}$  essentially in only a phase factor (if  $C_r$  and  $C_{r+\mu}$  are sufficiently distant from the ends of the grating for end effects to be negligible). That is,  $\chi_{r+\mu} \approx e^{i\mu ka \cos \alpha} \chi_r$ , the approximation becoming equality in the case of an infinite grating. The approximate solution of equations (3†) then involves sums which behave as

$$\sum e^{i|\mu|ka(1 \pm \cos \alpha)} / |\mu|^{\frac{1}{2}},$$

where the  $\pm$  sign is chosen accordingly as  $\mu \gtrless 0$ . Such sums may become very large as is the case, for example, when interaction effects are of importance. In particular, if either of  $ka(1 \pm \cos \alpha)$  is equal to an integral multiple of  $2\pi$ , then the sum is of order  $N^{\frac{1}{2}}$ . Either of the conditions

$$(5) \quad \begin{cases} ka(1 + \cos \alpha) = 2m\pi & m = 0, 1, 2, \dots, \\ ka(1 - \cos \alpha) = 2n\pi & n = 0, 1, 2, \dots, \end{cases}$$

defines one of the "Rayleigh wavelengths" mentioned previously.

It is apparent that if the Hankel functions in equations (4) could be replaced by functions which lead in the above manner to series with more rapidly diminishing terms (that is, with terms of order  $|\mu|^{-1-\epsilon}$ , where  $\epsilon > 0$ , so that when either of the conditions (5) obtains, the sum is of order  $\epsilon^{-1}$ ), and further, if the resulting integral equations could be solved approximately for small cylinders, then such a solution would in all probability be better suited to the discussion of interaction effects (and anomalous behavior) than the solution given in (†). Such functions may, in fact, be determined, and the resulting integral equations given a simple physical interpretation.

Suppose that  $(y_r, z_r)$ ,  $(y_\lambda, z_\lambda)$  represent a point of integration on  $C_r$ , and a fixed point on  $C_\lambda$ , respectively, where  $\nu \neq \lambda$ . Then, provided the cylinders are bounded in the  $z$ -direction, and  $ka|\nu - \lambda|$  is sufficiently large,

$$(6) \quad H(kr_{r\lambda}) \sim c(\nu - \lambda) e^{i\epsilon_{r\lambda} k(y_r - y_\lambda)},$$

where  $\epsilon_{r\lambda} = \pm 1$  according as  $\nu \gtrless \lambda$ , and

$$(7) \quad c(x) \equiv \left( \frac{2}{\pi ka} \right)^{\frac{1}{2}} e^{-\frac{1}{4}\pi i} \frac{e^{i|x|ka}}{|x|^{\frac{1}{2}}}, \quad x \neq 0.$$

In the light of the previous discussion it is therefore reasonable to subtract from each of equations (4) the troublesome series, and so obtain the following modified set of integral equations:

$$(8) \quad \sum \int_{C_\nu} G(\nu, \Lambda) \chi_\nu ds_\nu = -4iu^i(P_\Lambda) - e^{iky_\Lambda} \sum_{\nu < \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, 0) - e^{-iky_\Lambda} \sum_{\nu > \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, \pi),$$

where

$$(9) \quad G(\nu, \Lambda) \equiv \begin{cases} H(kr_{\nu\Lambda}) - c(\nu - \lambda) e^{i\epsilon_{\nu\Lambda} k(y_\nu - y_\Lambda)}, & \nu \neq \lambda, \\ H(kr_{\lambda\Lambda}), & \nu = \lambda, \end{cases}$$

and

$$(10) \quad \sigma_\nu(\alpha, \theta) \equiv \int_{C_\nu} \chi_\nu e^{-ik(y_\nu \cos \theta + z_\nu \sin \theta)} ds_\nu.$$

$\sigma_\nu(\alpha, \theta)$  is thus proportional to the far-field pattern of  $C_\nu$ , with  $\theta = 0$  corresponding to the positive  $y$ -axis.

It will be observed that  $G(\nu, \Lambda)$  is of order  $|\nu - \lambda|^{-3/2}$  when  $|\nu - \lambda| \gg 1$  and, when  $N$  is large, the series does not exhibit the undesirable behavior which was previously described. Furthermore, the driving terms on the right-hand side of equations (8) consist of the incident field, as in equations (4) and, in addition, terms corresponding to two plane waves propagating tangentially, and in opposite directions, along the grating. The amplitude and phase of the latter waves depend on the (unknown) field distribution on the contributing cylinders. The solution  $\chi_\nu$  of equations (8) will thus depend explicitly on the incident field and the interaction excitation embodied in the tangentially propagating plane wave terms.

An analysis with a similar physical interpretation has been employed by Karp (1955) and Karp and Radlow (1956) in connection with scattering by an infinite grating. In their approach, however, it is necessary to assume that the scatterers are separated by at least several wavelengths, and that each subtends only a small angle at a neighboring cylinder. This is in essence equivalent to the assumption, in equations (8), that  $G(\nu, \Lambda) = 0$  if  $\nu \neq \lambda$ , the conclusion drawn being that each element of the grating acts as though in isolation, excited only by the incident, and two tangentially propagating plane waves.

Equations (8), on the other hand, are rigorous, and involve no intrinsic restriction on separation, shape, or orientation of the cylinders. The assumption that  $G(\nu, \Lambda) \neq 0$  if  $\nu \neq \lambda$  is valid if the separation is sufficient, in which case the analysis is equivalent to that of Karp (1955) and Karp and Radlow (1956), and involves only the scattering characteristics of isolated cylinders. These properties are, in many cases, well known, or measurable, so the method of Karp can be applied with relative ease. The present procedure, which does not refer to the properties of cylinders in isolation except in the limit of large separation, is perhaps not so readily applicable. The chief advantage would

appear to lie in its validity no matter what the separation between, or orientation of the cylinders, although a solution of the integral equations seems to be possible only in the case of small scatterers.

While coefficients on the right-hand side of equations (8) are unknown, in that they depend on the functions  $\chi_\nu$ , the integral equations may be solved and the quantities  $\sigma_\nu(\alpha, 0)$ ,  $\sigma_\nu(\alpha, \pi)$  determined, at least formally, in the following manner.

Suppose that  $\xi_\lambda(\alpha)$  is the solution of

$$(11) \quad \sum_{C_\nu} G(\nu, \Lambda) \xi_\nu(\alpha) dS_\nu = u^1(P_\Lambda) = e^{i\lambda k a \cos \alpha} e^{ik(y_\Lambda \cos \alpha + z_\Lambda \sin \alpha)}.$$

Then since  $G(\nu, \Lambda)$  is independent of  $\alpha$ , the solution of equations (8) is

$$(12) \quad \chi_\lambda(\alpha) = -4i\xi_\lambda(\alpha) - e^{-i\lambda k a} \xi_\lambda(0) \sum_{\nu < \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, 0) - e^{i\lambda k a} \xi_\lambda(\pi) \sum_{\nu > \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, \pi),$$

in which the first term (which does include some interaction effects through  $G(\nu, \Lambda)$ ) arises from excitation by the incident plane wave, while the remaining terms are more directly attributable to interaction between the cylinders.

Multiplication of equations (12) by  $dS_\lambda \exp[-ik(y_\lambda \cos \theta + z_\lambda \sin \theta)]$ , and integration around  $C_\lambda$  determines  $\sigma_\lambda(\alpha, \theta)$  in the form

$$(13) \quad \sigma_\lambda(\alpha, \theta) = -4ip_\lambda(\alpha, \theta) - e^{-i\lambda k a} p_\lambda(0, \theta) \sum_{\nu < \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, 0) - e^{i\lambda k a} p_\lambda(\pi, \theta) \sum_{\nu > \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, \pi),$$

where

$$(14) \quad p_\lambda(\alpha, \theta) \equiv \int_{C_\lambda} \xi_\lambda(\alpha) e^{-ik(y_\lambda \cos \theta + z_\lambda \sin \theta)} dS_\lambda.$$

In equations (13), the functions  $p_\lambda(\alpha, \theta)$  are considered known. That is, they are determined by the solution of equations (11). From the physical point of view,  $p_\lambda(\alpha, \theta)$  is proportional to the far-field pattern of  $C_\lambda$  on which the wave function vanishes, and on which the prescribed distribution of  $\partial u / \partial n_\lambda$  is  $-4i\xi_\lambda(\alpha)$ ; but  $\xi_\lambda(\alpha)$  is not the solution for an isolated cylinder excited by  $u^1$  except when interaction is negligible or the cylinders are widely separated.

If, in equations (13),  $\theta$  is set to equal 0 and  $\pi$  in turn, then the following sets of linear algebraic equations are obtained for  $\sigma_\lambda(\alpha, 0)$  and  $\sigma_\lambda(\alpha, \pi)$ :

$$(15) \quad \left\{ \begin{array}{l} \sigma_\lambda(\alpha, 0) = -4ip_\lambda(\alpha, 0) - e^{-i\lambda k a} p_\lambda(0, 0) \sum_{\nu < \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, 0) \\ \quad - e^{i\lambda k a} p_\lambda(\pi, 0) \sum_{\nu > \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, \pi), \\ \sigma_\lambda(\alpha, \pi) = -4ip_\lambda(\alpha, \pi) - e^{-i\lambda k a} p_\lambda(0, \pi) \sum_{\nu < \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, 0) \\ \quad - e^{i\lambda k a} p_\lambda(\pi, \pi) \sum_{\nu > \lambda} c(\nu - \lambda) \sigma_\nu(\alpha, \pi). \end{array} \right.$$

The unknowns  $\partial u / \partial n_\lambda \equiv \chi_\lambda(\alpha)$  are therefore determined by equations (12), in terms of the solutions of equations (11) and (15). The rigorous solution of equations (11) for other than small values of  $N$  would appear to be a formidable task and, in general, some method of approximation must be employed. There is, however, one case for which an explicit solution may be obtained, namely the infinite grating of identical cylinders, and this will be considered in the following section.

### 3. THE INFINITE GRATING OF IDENTICAL CYLINDERS

#### 3.1. Solution of the Associated Linear Algebraic Equations

It has been mentioned previously that equation (1) is valid, whether the grating is finite or infinite. With regard to the latter case, the statement should, perhaps, be qualified to a certain extent. It is true if neither of the conditions (5) is satisfied. If one (or both) of these conditions obtains, then the application of Green's theorem may lead to terms in addition to those found in equations (1), because the integral around a large circle surrounding the scatterers does not, of necessity, tend to zero as the radius tends to infinity. As Karp and Radlow (1956) have pointed out, the solution at precisely the "critical frequencies" (Rayleigh wavelengths) defined by the conditions (5) may differ from the limit of the "non-critical" solution which would appear to be the solution of physical interest.

It may be concluded that the equations (15) are valid also for the infinite grating of identical cylinders in which the summation index runs from  $-\infty$  to  $+\infty$ , but periodicity of the structure brings a reduction to two equations in two unknowns ( $\sigma_0(\alpha, 0)$  and  $\sigma_0(\alpha, \pi)$ , say). For, since

$$(16) \quad [\chi_\lambda(\alpha), \xi_\lambda(\alpha)] = e^{i\lambda ka \cos \alpha} [\chi_0(\alpha), \xi_0(\alpha)],$$

the equations (15) for  $\lambda = 0$  have the solution

$$(17) \quad \begin{cases} \sigma(\alpha, 0) = 4i \frac{p(\alpha, \pi)p(\pi, 0)\sum_- - p(\alpha, 0)[1 + p(\pi, \pi)\sum_+]}{[1 + p(0, 0)\sum_-][1 + p(\pi, \pi)\sum_+] - p(\pi, 0)p(0, \pi)\sum_+\sum_-} \\ \sigma(\alpha, \pi) = 4i \frac{p(\alpha, 0)p(0, \pi)\sum_- - p(\alpha, \pi)[1 + p(0, 0)\sum_-]}{[1 + p(0, 0)\sum_-][1 + p(\pi, \pi)\sum_+] - p(\pi, 0)p(0, \pi)\sum_+\sum_-} \end{cases}$$

(in which it is assumed that the denominators do not vanish). In equations (17) and in the following analysis, where feasible, the subscript "0" has been omitted, while

$$(18) \quad \sum_{\pm} \equiv \sum_{\nu=1}^{\infty} c(\nu) e^{\pm i\nu ka \cos \alpha}.$$

The function  $\chi(\alpha) [\equiv \chi_0(\alpha)]$  is then determined by equation (12), with  $\lambda = 0$ :

$$(19) \quad \chi(\alpha) = -4i\xi(\alpha) - \xi(0)\sigma(\alpha, 0)\sum_- - \xi(\pi)\sigma(\alpha, \pi)\sum_+.$$

To one familiar with the work of Karp (1955) and Karp and Radlow (1956) the similarity in form between equation (19) and a corresponding expression obtained by these authors is apparent. The difference lies in the fact that  $\xi$

is not the solution for scattering by an isolated cylinder, but is asymptotic to such a function when interaction becomes negligible. When interaction is of importance, then one (or both) of  $\sum_{\pm}$  becomes very large, and is unbounded at Rayleigh wavelengths. Nevertheless the individual terms in equation (19) remain finite, the limiting forms being readily obtained with the aid of equations (17).

The solution of the boundary value problem under consideration has thus been reduced to the determination of the function  $\xi(\alpha)$ . This in turn is given as the solution of equation (11) when specialized to the case of the infinite grating. However, before attempting to find an approximate solution to this equation, it is worth while to consider in some detail the field of the grating.

### 3.2. The Field of the Grating

The quantity which is possibly of most physical interest (and is certainly the most readily observable) is the field distribution in the space away from the grating elements. The field at a point  $P(y, z)$  not on any of the cylinders is determined by equation (1), in which  $\nu$  runs from  $-\infty$  to  $+\infty$ . Periodicity of the structure permits all integrations to be performed around  $C_0$ , and

$$(20) \quad u(P) = u^1(P) + \frac{1}{4}i \int_{C_0} ds \left( u \frac{\partial}{\partial n} - \frac{\partial u}{\partial n} \right) \sum e^{i \nu k a \cos \alpha} H(kR_{\nu}),$$

where

$$(21) \quad R_{\nu}^2 = [y - (\nu a + y_0)]^2 + [z - z_0]^2.$$

The infinite series appearing in equation (20) may be written in an alternative and more rapidly convergent form by an application of the Poisson summation formula [see, for example, Wait 1955, equation (12)]:

$$(22) \quad \sum_{\nu=-\infty}^{\infty} e^{i \nu k a \cos \alpha} H(kR_{\nu}) = 2 \sum_{\nu=-\infty}^{\infty} \frac{e^{i(k \cos \alpha + 2\pi\nu/a)(y-y_0) + i\kappa_{\nu}|z-z_0|}}{\kappa_{\nu} a},$$

where

$$(23) \quad \kappa_{\nu} = [k^2 - (k \cos \alpha + 2\pi\nu/a)^2]^{\frac{1}{2}}$$

$$= \begin{cases} |\kappa_{\nu}|, & k > |k \cos \alpha + 2\pi\nu/a|, \\ i|\kappa_{\nu}|, & 0 < k < |k \cos \alpha + 2\pi\nu/a|. \end{cases}$$

The vanishing of  $\kappa_{\nu}$  for some  $\nu$  implies the satisfaction of at least one of the conditions (5) which define the Rayleigh wavelengths.

The total field is therefore of the form

$$(24) \quad u(P) = u^1(P) + \frac{1}{2}i \int_{C_0} ds \left( u \frac{\partial}{\partial n} - \frac{\partial u}{\partial n} \right) \sum \frac{e^{i(k \cos \alpha + 2\pi\nu/a)(y-y_0) + i\kappa_{\nu}|z-z_0|}}{\kappa_{\nu} a}.$$

Equation (24) exhibits clearly the plane wave spectral character of the field if the point  $(y, z)$  does not lie between elements of the grating, whatever the boundary condition on the cylinders. For the particular boundary value being considered at present, and when  $P$  is outside a strip containing the cylinders (so  $z \neq z_0$  for any point on  $C_0$ ), the field is of the form

$$(25) \quad u(P) = u^1(P) + \sum a_\nu^\pm e^{ik(y \cos \theta_\nu \pm z \sin \theta_\nu)},$$

the  $\pm$  sign being chosen accordingly as  $z \gtrless z_0$ . Here

$$(26) \quad a_\nu^\pm = -\frac{1}{2}i\sigma(\alpha, \pm\theta_\nu)/(\kappa_\nu a),$$

where  $\sigma$  is determined by equation (13), with  $\lambda = 0$ , the (possibly complex) angle  $\theta$ , is defined by

$$(27) \quad \begin{cases} k \cos \theta_\nu = k \cos \alpha + 2\pi\nu/a, \\ k \sin \theta_\nu = \kappa_\nu \end{cases}$$

and

$$0 \leq \operatorname{Re} \theta_\nu \leq \pi.$$

It may be noted that, although  $\sigma(\alpha, \pi) = \sigma(\alpha, -\pi)$ , in general  $\sigma(\alpha, \theta_\nu) \neq \sigma(\alpha, -\theta_\nu)$ , so  $a_\nu^+ \neq a_\nu^-$ . It may also be verified that  $\sigma(\alpha, \pm\theta_\nu) \rightarrow 0$  as  $\kappa_\nu \rightarrow 0$ , and the  $a_\nu^\pm$  possess finite limits. These are found perhaps most readily by employing the integral form of  $\sigma(\alpha, \pm\theta_\nu)$  (equation (10)) in equation (26). If  $\kappa_\nu \rightarrow 0$  for  $\nu = n > 0$ , say, then  $\theta_n \rightarrow 0$ , the corresponding wave is propagated tangentially along the grating, and

$$(28) \quad \lim a_n^\pm = -\frac{1}{2}i \lim \left[ \frac{\sigma(\alpha, 0)}{\kappa_n a} \pm \frac{1}{ka} \frac{\partial}{\partial \theta} \sigma(\alpha, \theta) \Big|_{\theta=0} \right].$$

(This expression is valid whether the limit is approached by variation of  $\alpha$ ,  $ka$ , or both, about the critical values. Explicit differentiation with respect to  $\alpha$  is not required since  $a$  tends to its limiting value more quickly than  $\kappa_n^{-\frac{1}{2}}$  tends to zero.) The first term in the square bracket is indeterminate for  $\kappa_n = 0$ , for  $|\sum_-|$  approaches infinity (see equations (17)). Karp and Radlow (1956) have noted that, if  $ka(1 \pm \cos \alpha) = 2m\pi + \delta_\pm$  where  $m$  is integral, then  $\sum_\pm \rightarrow [2/(ka\delta_\pm)]^{\frac{1}{2}}$  as  $\delta_\pm \rightarrow 0$ , where the root  $\delta_\pm^{\frac{1}{2}}$  is positive, or positive imaginary (see also the Appendix to the present paper). Then  $\kappa_n a = \delta_-^{\frac{1}{2}}(2ka - \delta_-)^{\frac{1}{2}}$ , and  $\kappa_n a \sum_- \rightarrow 2$ . Thus

$$(29) \quad \lim \frac{\sigma(\alpha, 0)}{\kappa_n a} = 2i \left\{ p(\alpha, \pi)p(\pi, 0) \sum_+ - p(\alpha, 0) \left[ 1 + p(\pi, \pi) \sum_+ \right] \right\} / D,$$

where

$$(30) \quad D = p(0, 0)[1 + p(\pi, \pi)\sum_+] - p(\pi, 0)p(0, \pi)\sum_+,$$

while the appropriate limiting form of  $\partial\sigma/\partial\theta$  appearing in equation (28) may be found with the aid of equations (13), (14), and (17).

If, on the other hand,  $\kappa_\nu \rightarrow 0$  for  $\nu = -m < 0$ , then  $\theta_{-m} \rightarrow \pi$ ,  $|\sum_+| \rightarrow \infty$ , and

$$(31) \quad \lim a_{-m}^\pm = -\frac{1}{2}i \lim \left[ \frac{\sigma(\alpha, \pi)}{\kappa_{-m} a} \mp \frac{1}{ka} \frac{\partial}{\partial \theta} \sigma(\alpha, \theta) \Big|_{\theta=\pi} \right].$$

The limit of  $\sigma(\alpha, \pi)/(\kappa_{-m} a)$  may be found from the right-hand side of equation (29) by interchanging  $\pi$  and  $0$ , and replacing  $\sum_+$  by  $\sum_-$ .

In the case of a "double anomaly", when both  $\kappa_n, \kappa_{-m} \rightarrow 0$  simultaneously, the appropriate limits are readily found. For instance, the limit of  $\sigma(\alpha, 0) / (\kappa_n a)$  is then equal to the limit, as  $|\sum_{\pm}| \rightarrow \infty$  of the right-hand side of equation (29).

The behavior at a Rayleigh wavelength of the amplitudes of the remaining (non-grazing) waves may be noted. For example, if  $\kappa_n \rightarrow 0, n > 0$ , then, if  $\nu \neq n$

$$(32) \quad a_{\nu}^{\pm} \rightarrow -\frac{2}{\kappa_n a} \left\{ p(\alpha, \pm\theta_{\nu}) \right. \\ \left. + p(0, \pm\theta_{\nu}) \left( [p(\alpha, \pi)p(\pi, 0) - p(\alpha, 0)p(\pi, \pi)] \sum_{\pm} - p(\alpha, 0) \right) / D \right. \\ \left. + p(\pi, \pm\theta_{\nu}) [p(\alpha, 0)p(0, \pi) - p(\alpha, \pi)p(0, 0)] \sum_{\pm} / D \right\}.$$

This limit is similar in form to the corresponding expression found by Karp and Radlow (1956), but because the  $p$ -functions do not pertain to a cylinder in isolation, the two results are not identical.

A thorough examination of the behavior of the amplitudes  $a_{\nu}^{\pm}$  for a grating of identical cylinders of arbitrary form appears to be out of the question, because knowledge of the function  $\xi(\alpha)$  is required. Nevertheless, some idea of the possible variation of  $a_{\nu}^{\pm}$  may be obtained (even though an explicit solution  $\xi(\alpha)$  is unknown) by assuming that, in the regions of interest, all functions with the possible exception of  $\sum_{\pm}$  are slowly varying. (This appears to be a legitimate assumption, since the kernel of the integral equation satisfied by  $\xi(\alpha)$  has been chosen essentially with this point in view.) In particular, the behavior near the Rayleigh wavelengths may be discussed, and attention will be confined to this region for the remainder of the present section.

It is apparent that a number of special cases needs consideration. With reference to the amplitudes  $a_{\nu}^{\pm}, \nu \neq n, -m$ , the following situations arise:

- (i)  $\kappa_n \rightarrow 0, n > 0; n \neq \nu; \theta_{\nu} \neq 0, \pi.$
- (ii)  $\kappa_{-m} \rightarrow 0, m > 0; m \neq \nu; \theta_{\nu} \neq 0, \pi.$
- (iii)  $\kappa_n, \kappa_{-m} \rightarrow 0$  simultaneously;  $n, m \neq \nu; \theta_{\nu} \neq 0, \pi.$

(i) and (ii) are essentially equivalent, and only the former will be considered in any detail. (iii) corresponds to a "double anomaly", in which case both conditions (5) are satisfied for some pair of integers  $m, n$ . In these circumstances, it is necessary that  $\tan^2 \frac{1}{2}\alpha = n/m$ , and  $ka = (n+m)\pi$ .

For the waves which propagate tangentially at the Rayleigh wavelengths, the following cases apply:

- (iv)  $\kappa_n \rightarrow 0, n > 0; \theta_n \rightarrow 0.$
- (v)  $\kappa_{-m} \rightarrow 0, m > 0; \theta_{-m} \rightarrow \pi.$
- (vi)  $\kappa_n, \kappa_{-m} \rightarrow 0; \theta_n \rightarrow 0, \theta_{-m} \rightarrow \pi$ , simultaneously.

Here (iv) and (v) are essentially equivalent.

It will be recalled (equation (13)) that, if interaction is small and  $\Sigma_{\pm}$  negligible, then  $\sigma(\alpha, \theta) \doteq -4ip(\alpha, \theta)$ . Also  $-4i\xi(\alpha)$  approximates to the distribution of  $\partial u / \partial n$  on an isolated cylinder excited by  $u^1$ , so  $p(\alpha, \theta)$  is, for  $ka$  sufficiently large, other than in the neighborhood of the Rayleigh wavelengths, approximately proportional to the far-field pattern of the isolated cylinder  $C_0$  excited by  $u^1$ . Near the critical wavelengths,  $p(\alpha, \theta)$  departs from the polar diagram of an isolated cylinder, but may be assumed to be slowly varying. Since the purpose of the present investigation is the study of anomalous effects, it is reasonable to discuss the behavior of the "normalized" function  $(\kappa_r a)^{-1}\omega(\alpha, \pm\theta_r)$ , where

$$(33) \quad -4ip(\alpha, \theta)\omega(\alpha, \theta) = \sigma(\alpha, \theta),$$

and it is assumed that  $p(\alpha, \theta) \neq 0$  in the interval under consideration.

An explicit expression for  $\omega(\alpha, \theta_r)$  may be obtained with the aid of equations (13) and (17). It is seen that  $\omega(\alpha, \theta_r)$  may be written as

$$(34) \quad \omega(\alpha, \theta_r) = \frac{1+X_1\Sigma_{+}+X_2\Sigma_{-}+X_3\Sigma_{+}\Sigma_{-}}{1+Y_1\Sigma_{+}+Y_2\Sigma_{-}+Y_3\Sigma_{+}\Sigma_{-}},$$

where  $X_n, Y_n$  are complex, slowly varying functions of  $\alpha, ka$ . If the quantities  $\delta_{\pm}$  (at least one of which is assumed to be small) are defined by

$$(35) \quad ka(1 \pm \cos \alpha) = 2 \left( \frac{m}{n} \right) \pi + \delta_{\pm},$$

where one or other of  $\delta_{\pm}$  vanishes at a Rayleigh wavelength, then

$$(36) \quad \Sigma_{\pm} = \left( \frac{2}{\pi ka} \right)^{\frac{1}{2}} e^{-\frac{1}{4}\pi i} \sum_{v=1}^{\infty} \frac{e^{iv\delta_{\pm}}}{v^{\frac{1}{2}}}.$$

The sum appearing in equation (36) may be transformed into a more rapidly convergent series by an application of the Poisson summation formula (see Appendix). Then, for sufficiently small  $|\delta_{\pm}|$ ,

$$(37) \quad \Sigma_{\pm} \doteq \left( \frac{2}{\pi ka} \right)^{\frac{1}{2}} e^{-\frac{1}{4}\pi i} \begin{cases} X(|\delta_{\pm}|) + iY(|\delta_{\pm}|), & \delta_{\pm} > 0, \\ X(|\delta_{\pm}|) - iY(|\delta_{\pm}|), & \delta_{\pm} < 0, \end{cases}$$

where

$$(38) \quad X(z) + iY(z) = \left( \frac{2\pi}{z} \right)^{\frac{1}{2}} \left[ \frac{e^{\frac{1}{4}\pi i}}{\sqrt{2}} - C(z) - iS(z) \right], \quad z > 0,$$

$$(39) \quad C(z) + iS(z) = \frac{1}{\sqrt{(2\pi)}} \int_0^z \frac{e^{it}}{t^{\frac{1}{2}}} dt,$$

and the functions  $C(z), S(z)$  have been tabulated by Jahnke and Emde (1945, p. 35). The approximation (37) for  $\Sigma_{\pm}$  may also be obtained by an application of the Euler-Maclaurin formula (see, for example, Jeffreys and Jeffreys 1956, p. 278).

Since  $\Sigma_{\pm}$  are now expressed in terms of well-known functions, it is possible to give some consideration to the six cases listed above.

(i) Here  $|\Sigma_-| \gg 1$ , while  $\Sigma_+$  is relatively small and slowly varying.  $\Sigma_-$  may be approximated by the expression (37). Then  $\omega(\alpha, \theta_s)$  is of the form

$$(40) \quad \omega(\alpha, \theta_s) = \frac{1 + (a_1 + ib_1) + (a_2 + ib_2)(X \pm iY)}{1 + (c_1 + id_1) + (c_2 + id_2)(X \pm iY)},$$

where, for sufficiently large values of  $ka$ , the conditions  $a_n^2 + b_n^2 \ll 1$ ,  $c_n^2 + d_n^2 \ll 1$ , obtain. The argument of the functions  $X, Y$  is  $|\delta_-|$ , the  $\pm$  sign being chosen accordingly as  $\delta_- \gtrless 0$ . The quantities  $a_n, b_n, c_n$ , and  $d_n$  depend on the  $p$ -functions, and without knowledge of  $\xi(\alpha)$ , one can only hazard a guess as to their values. If second-order terms in small quantities are, in general, neglected, then

$$(41) \quad |\omega(\alpha, \theta_s)|^2 \propto \frac{1 + 2(a_2 X \mp b_2 Y) + (a_2^2 + b_2^2)(X^2 + Y^2)}{1 + 2(c_2 X \mp d_2 Y) + (c_2^2 + d_2^2)(X^2 + Y^2)}, \quad \delta_- \gtrless 0$$

$= I$ , say.

The slowly varying factor  $(\kappa_s a)^{-1}$  has been omitted from consideration in the previous discussion.

(ii) The behavior of  $|\omega|^2$  is, in this case, similar to that outlined above for (i), with  $\delta_-$  replaced by  $\delta_+$ .

(iii) Here both  $|\Sigma_{\pm}| \gg 1$ , and each is approximated by the expression (37). It may be shown that, if  $\alpha$  is fixed so that  $\tan^2 \frac{1}{2}\alpha = n/m$ , while  $ka$  is varied, then  $\delta_- = \delta_+ \tan^2 \frac{1}{2}\alpha$ , and  $\delta_{\pm}$  have the same sign. On the other hand, if  $ka$  is fixed, while  $\alpha$  is varied, then  $\delta_+ + \delta_- = 0$ , and  $\delta_{\pm}$  have opposite signs. In either case, the possible variations in  $|\omega|^2$  near  $\delta_{\pm} = 0$  are similar. The expression for  $\omega$  is considerably more complex than (40), but, if  $\delta_{\pm}$  are sufficiently small that  $C(|\delta_{\pm}|)$  and  $S(|\delta_{\pm}|)$  may be neglected, then

$$(42) \quad |\omega(\alpha, \theta_s)|^2 \propto \begin{cases} \frac{e_1 |\delta_-|^{\frac{1}{2}} + f}{g_1 |\delta_-|^{\frac{1}{2}} + h}, & \delta_- > 0, \\ \frac{e_2 |\delta_-|^{\frac{1}{2}} + f}{g_2 |\delta_-|^{\frac{1}{2}} + h}, & \delta_- < 0, \end{cases}$$

where the coefficients are real and  $f > 0, h > 0$ .

The expressions for  $|\omega|^2$  in (i) and (ii) are also of this form when  $|\delta_-|, |\delta_+|$ , respectively, are sufficiently small. The derivative of  $|\omega|^2$  (equation (42)) with respect to  $\delta_-$  is  $\propto \text{sgn}(e_1 h - g_1 f)$  for  $\delta_- = 0+$ , and  $-\propto \text{sgn}(e_2 h - g_2 f)$  for  $\delta_- = 0-$ , provided  $e_1 h \neq g_1 f, e_2 h \neq g_2 f$ . If equality obtains in either case, the corresponding derivative is finite. Evidently many different types of behavior of  $|\omega|^2$  are possible near a Rayleigh wavelength.

In Fig. 2 are plotted, as functions of  $\delta_-$ , typical curves of  $I$  (which is proportional to the spectral intensity) calculated from the expression (41), for a variety of values of the coefficients. A change of sign in  $b_2, d_2$  leads to reflection in the line  $\delta_- = 0$ . Similar behavior is exhibited by the expression (42). Whether all these variations may in fact arise is not known, because the complexity of the coefficients makes it difficult, if not impossible, to reach any firm conclusions about the signs or magnitudes when the cylinder shape and spacing are arbitrary.

(iv) The study of case (iv) requires consideration of  $\omega(\alpha, \theta_n)/(\kappa_n a)$  as  $\kappa_n \rightarrow 0$  and  $\theta_n \rightarrow 0$ . With reference to the numerator in equation (34), it is found that the coefficient of  $\sum_-$  is of order  $\theta_n$  (that is, of order  $\kappa_n$ , or  $\delta_-^{-\frac{1}{2}}(2ka - \delta_-)^{\frac{1}{2}}$ ) as  $\theta_n \rightarrow 0$ . Thus

$$\left( \frac{\omega(\alpha, \theta_n)}{\kappa_n a} \right)^2 \propto \frac{1+2(tC \mp uS) + (t^2 + u^2)(C^2 + S^2)}{|\delta_-| [1+2(c_2 X \mp d_2 Y) + (c_2^2 + d_2^2)(X^2 + Y^2)]}, \quad \delta_- \gtrless 0, \\ = I', \text{ say,}$$

where the argument of the functions is  $|\delta_-|$ ,  $t$ , and  $u$  being slowly varying. In equation (43), it is assumed that  $t^2 + u^2 \ll 1$ ,  $c_2^2 + d_2^2 \ll 1$  and, as in equation (41), some second-order terms in small quantities have been neglected.

(v) Case (v) implies similar behavior near a Rayleigh wavelength.

(vi) In the immediate neighborhood of a double anomaly, it is found that  $|\omega(\alpha, \theta_n)/(\kappa_n a)|^2$  and  $|\omega(\alpha, \theta_{-m})/(\kappa_{-m} a)|^2$  behave very much like the expressions (42).

Typical curves of  $I'$  as a function of  $\delta_-$  for several values of  $t$ ,  $u$ ,  $c_2$ , and  $d_2$  (expression (43)) are shown in Fig. 3. These, too, indicate a rapid variation in intensity near  $\delta_- = 0$ . In Figs. 3(a), (b), and (c), the slope is infinite at  $\delta_- = 0$ ; in Fig. 3(d), the slope is finite as  $\delta_- \rightarrow 0+$ , and infinite as  $\delta_- \rightarrow 0-$ .

The foregoing discussion should give some idea of possible behavior of the wave amplitudes near a Rayleigh wavelength for a grating with elements of general shape and separation sufficiently large that certain approximations are justified. In order to proceed further, some knowledge of  $\xi(\alpha)$  is essential, and in the following section, approximations to the solution of the integral equation and field distribution are obtained for an infinite grating of elliptic cylinders, the cross-sectional dimensions of which are small in comparison with the wavelength.

### 3.3. An Approximate Solution for Small Elliptic Cylinders

#### 3.3.1. The Determination of $\chi(\alpha)$

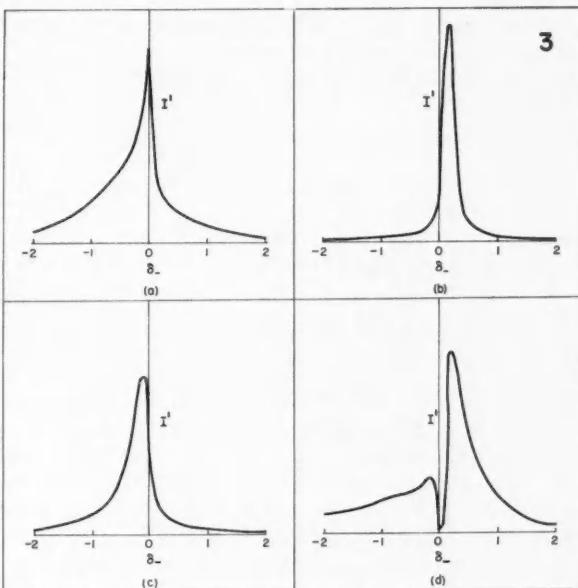
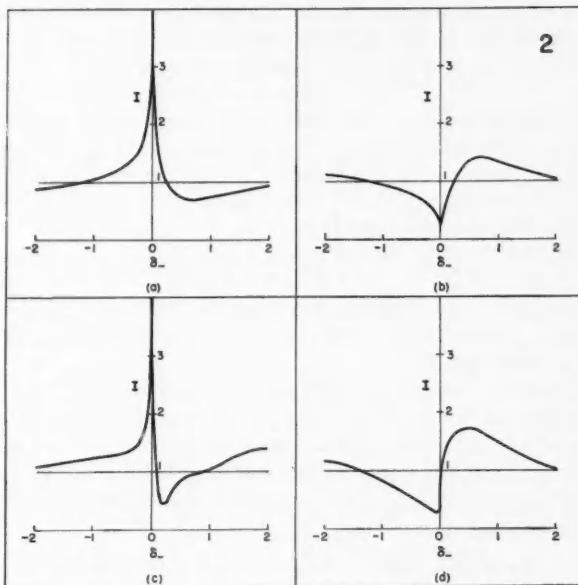
A set of integral equations for the functions  $\xi_r(\alpha)$  (equations (11)) has been derived in a previous section. For the infinite grating,  $\xi_r = \xi_0 e^{i_r k a \cos \alpha}$ ,

FIG. 2. Examples of (theoretically) possible variations in intensity of non-grazing waves near a Rayleigh wavelength for a grating on which the total wave function is null. Values of the parameters in equation (41) are given below.

	(a)	(b)	(c)	(d)
$a_2$	0.2	0.1	-0.2	0.1
$b_2$	0.2	0.1	0.2	-0.1
$c_2$	0.1	0.2	-0.1	0.2
$d_2$	0.1	0.2	0.1	0.1

FIG. 3. Examples of (theoretically) possible variations in intensity of grazing waves near a Rayleigh wavelength for a grating on which the total wave function is null. Values of the parameters in equation (43) are given below.

	(a)	(b)	(c)	(d)
$t$	-0.25	-0.25	-0.25	-0.25
$u$	0.5	0.5	0.5	0.5
$c_2$	0.2	-0.1	0.1	0.2
$d_2$	-0.1	0.2	-0.2	0.2



FIGS. 2-3.

the set reduces to one equation in the unknown function  $\xi(\alpha)[\equiv \xi_0(\alpha)]$ , and the integrations may all be performed around the contour  $C_0$ .

$$(44) \quad \int_{C_0} ds \xi \sum G_\nu e^{i k a \cos \nu \alpha} = u^1(P_0),$$

where the subscript "0" will now imply a point fixed on  $C_0$ , and

$$(45) \quad G_\nu = \begin{cases} H(kr_{\nu 0}) - c(\nu) e^{\pm ik(y - y_0)}, & \nu \neq 0, \\ H(kr_{00}), & \nu = 0. \end{cases}$$

The distance  $r_{\nu 0}$  is, for all  $\nu$ , given by

$$(47) \quad r_{\nu 0}^2 = [y_0 - (\nu a + y)]^2 + [z_0 - z]^2.$$

The infinite series which defines the kernel of equation (44) is convergent for all angles of incidence  $\alpha$ , and  $ka > 0$ , including such as satisfy either or both of the conditions (5).

An approximate analytic solution of equation (44) for elliptic cylinders may be found subject to the restriction that the cross-sectional dimensions of the cylinders are very much less than the wavelength. If it is assumed that the major axes are all inclined at an angle  $\beta$  to the positive  $y$ -axis (Fig. 1), then the point of integration  $(y, z)$  on  $C_0$  has elliptic co-ordinates  $(\mu, \eta)$ , where  $\mu$  is constant on  $C_0$ , and

$$(48) \quad \begin{cases} y = b(\cosh \mu \cos \beta \cos \eta - \sinh \mu \sin \beta \sin \eta) \equiv bY, \\ z = b(\cosh \mu \sin \beta \cos \eta + \sinh \mu \cos \beta \sin \eta) \equiv bZ. \end{cases}$$

Here the normalized co-ordinates  $(Y, Z)$  have been introduced, and  $2b$  is the interfocal distance. The fixed point  $P_0$  has co-ordinates  $(\mu, \eta_0)$  in the elliptic system, and  $(Y_0, Z_0)$  in the normalized system.

If it is assumed that expansions exist of the form

$$(49) \quad \begin{cases} b\xi(\alpha)(\cosh^2 \mu - \cos^2 \eta)^{\frac{1}{2}} = \sum_{n=0}^{\infty} \xi_n(\alpha) \epsilon^n, \\ u^1(P_0) = \sum_{n=0}^{\infty} u_n^1(P_0) \epsilon^n, \\ \sum G_\nu e^{i k a \cos \nu \alpha} = \sum_{n=0}^{\infty} g_n \epsilon^n, \end{cases}$$

where  $\epsilon = kb \ll 1$ , while  $\xi_n(\alpha)$  (not to be confused with  $\xi_\nu(\alpha)$ ; this symbol was used previously to denote the value of  $\xi$  on  $C_\nu$ ) and  $g_n$  depend on  $\log \epsilon$ , then  $\xi_0, \xi_1, \dots$ , satisfy the integral equations

$$(50) \quad \begin{cases} \int_0^{2\pi} \xi_0 g_0 d\eta = u_0^1(P_0), \\ \int_0^{2\pi} \xi_1 g_0 d\eta = u_1^1(P_0) - \int_0^{2\pi} \xi_0 g_1 d\eta, \\ \vdots \end{cases}$$

Here the relationship  $ds = b(\cosh^2 \mu - \cos^2 \eta)^{1/2} d\eta$  has been introduced. (Perhaps it should also be pointed out that the expansion of  $b\xi(\alpha)(\cosh^2 \mu - \cos^2 \eta)^{1/2}$  is, strictly speaking, in powers of  $\epsilon e^\mu$ , where  $2be^\mu$  is equal to the sum of the lengths of the major and minor axes of the ellipse.)

The solution of equations (50) is obtained in a straightforward manner, the method having been employed in (†), and details will be omitted. The second and third of expansions (49) may be found explicitly by reference to equations (5†), (6†), (7†), (16†), and (17†), together with equation (45), and it may be verified that the expansion in powers of  $\epsilon$  of  $G_\nu$ ,  $\nu \neq 0$ , is convergent if  $|\nu|a > 2b \cosh \mu$ . Thus, in order that the series  $\sum g_n \epsilon^n$  be meaningful, it is necessary that  $a > 2b \cosh \mu$  (that is,  $a$  be greater than the length of the major axis), and this imposes a lower limit on the separation  $a$  for any given  $b$  and  $\mu$ . Then,

$$(51) \quad g_0 = \frac{2i}{\pi} \psi_0 + \sum_0(\alpha, ka),$$

where

$$(52) \quad \begin{cases} \psi_0 = p + \mu - 2 \sum_{n=1}^{\infty} \frac{e^{-n\mu}}{n} (\cosh n\mu \cos n\eta \cos n\eta_0 + \sinh n\mu \sin n\eta \sin n\eta_0), \\ p = \log(\frac{1}{4}\gamma\epsilon) - \frac{1}{2}\pi i, \\ \log \gamma = 0.5772157 \dots \end{cases} \quad (\text{Euler's constant}),$$

and

$$(53) \quad \sum_0(\alpha, ka) \equiv \sum' [H(|\nu|ka) - c(\nu)] e^{i\nu ka \cos \alpha},$$

the prime implying the omission of the term for  $\nu = 0$ .

Because  $u_0^t(P_0) = 1$ ,  $\xi_0(\alpha)$  is a constant with respect to position on the cylinder, determined by

$$(54) \quad 2\pi\xi_0(\alpha) = \left[ \frac{2i}{\pi} (p + \mu) + \sum_0(\alpha, ka) \right]^{-1},$$

where the terms in the infinite series are of order  $|\nu|^{-3/2}$  for  $|\nu| \gg 1$ . Thus  $\xi_0(\alpha)$  does not vanish for any  $ka$  and  $\alpha$ . Also  $\xi_0(0) = \xi_0(\pi) \equiv \xi_0$ , say.

The determination of the corresponding approximation to  $\chi(\alpha)$  from equation (19) necessitates the evaluation of  $\sigma^{(0)}(\alpha, 0)$  and  $\sigma^{(0)}(\alpha, \pi)$ . These quantities may be defined by equations (17) with the  $p$ -functions replaced by the first term  $p_0$  in their series expansions. In this connection, a few words about notation are in order. A superscript "n", such as in  $\sigma^{(n)}(\alpha, 0)$ , implies the  $n$ th-order approximation, while a subscript "n", as in  $\xi_n(\alpha)$  denotes the coefficient of  $\epsilon^n$  in a series expansion. Here it will be assumed that the angle  $\theta$  is real, or has at most a small imaginary part; thus  $\cos \theta$  and  $\sin \theta$  are small, and the exponential factor in equation (14) may be expanded in powers of  $\epsilon$ , the coefficients of each power being small. If, however,  $\theta$  were to have a greater imaginary part, then these coefficients might be large. In such a case, the exponential should not be expanded, and the resultant integration would

lead to Bessel functions. In particular, some of the following results may not be valid for the amplitudes of highly evanescent waves.

According to equation (14),  $p_0(\alpha, \theta)$  is defined as

$$(55) \quad p_0(\alpha, \theta) = 2\pi\xi_0(\alpha),$$

which is independent of  $\theta$ . Thus  $\sigma^{(0)}(\alpha, \theta)$  (equation (13)) is independent of  $\theta$ , and equal to  $\sigma^{(0)}(\alpha)$ , say.

$$(56) \quad \sigma^{(0)}(\alpha) = -8\pi i \frac{\xi_0(\alpha)}{1+2\pi\xi_0(\sum_+ + \sum_-)}$$

and equation (19) gives, as the zero-order approximation to  $\chi(\alpha)ds/d\eta$ :

$$(57) \quad \chi^{(0)}(\alpha) \frac{ds}{d\eta} = -4i \frac{\xi_0(\alpha)}{1+2\pi\xi_0(\sum_+ + \sum_-)}.$$

It is evident that  $\chi^{(0)}(\alpha)$  tends to zero when either (or both) of  $|\sum_{\pm}|$  tends to infinity, in agreement with the usual expression for current on an infinite grating of small, perfectly conducting wires (see, for example, Wait 1955). This agreement, however, appears to be fortuitous, for if the current density ( $I(\alpha)$ , say) is defined as the first term in the formal solution, for an infinite grating of small wires, of equation (4), then it is not difficult to show that

$$(58) \quad I(\alpha) \frac{ds}{d\eta} = -4i \frac{\xi_0(\alpha)}{1+2\pi\xi_0(\alpha)(\sum_+ + \sum_-)}.$$

This expression for the current density, which is essentially the form derived several times in the past, is equal to  $\chi^{(0)}(\alpha)$  only when  $\sum_{\pm}$  are negligible, or infinite.\*

The function  $\chi^{(0)}(\alpha)$  tends to zero at a Rayleigh wavelength. In such a case, it is of interest to consider the next approximation, and to determine  $\xi_1(\alpha)$  from the second of equations (50). The function  $g_1$  may be written in the form

$$(59) \quad g_1 = (Y_0 - Y) \sum_1(\alpha, ka),$$

where

$$(60) \quad \sum_1(\alpha, ka) \equiv \sum' \epsilon_{\nu 0} [H_1^{(1)}(|\nu|ka) + iC(\nu)] e^{i\nu ka \cos \alpha},$$

$\epsilon_{\nu 0} = \pm 1$  as  $\nu \gtrless 0$ , and the terms in the sum are, as expected, of order  $|\nu|^{-3/2}$  when  $|\nu| \gg 1$ . Because

$$(61) \quad u_1^1(P_0) = i[\cosh \mu \cos(\alpha - \beta) \cos \eta_0 + \sinh \mu \sin(\alpha - \beta) \sin \eta_0],$$

it is not difficult to see that

$$(62) \quad 2\pi\xi_1(\alpha) = -ie^{\mu} \cos(\alpha - \beta - \eta_0) + 2\pi e^{\mu} \xi_0(\alpha) \sum_1(\alpha, ka) \cos(\beta + \eta_0).$$

\*It is not intended to imply that  $\chi^{(0)}(\alpha)$  is the unique first approximation, and that  $I(\alpha)$  is incorrect, for another (unpublished) small cylinder solution obtained by a different method, and in which higher-order terms are small, gives  $I(\alpha)$  as the first approximation. It would appear then that there is no unique small cylinder solution to the problem. Nevertheless, the two expressions are asymptotically equal in the limit as  $\epsilon$  tends to zero.

Equation (14) then implies that

$$(63) \quad p_1(\alpha, \theta) = 0$$

for all  $\alpha, \theta$  so

$$(64) \quad \sigma^{(1)}(\alpha, \theta) \equiv \sigma^{(0)}(\alpha)$$

and equation (19) gives, for  $\chi^{(1)}(\alpha)$ ,

$$(65) \quad \chi^{(1)}(\alpha) \frac{ds}{d\eta} = \chi^{(0)}(\alpha) \frac{ds}{d\eta} - 4i\epsilon \left[ \xi_1(\alpha) - 2\pi\xi_0(\alpha) \frac{\xi_1(\pi) \sum_+ + \xi_1(0) \sum_-}{1 + 2\pi\xi_0(\sum_+ + \sum_-)} \right].$$

Some algebraic manipulation, and use of the conditions (5), yields the following limiting forms of  $\chi^{(1)}(\alpha)$ :

$$(66) \quad \chi^{(1)}(\alpha) \frac{ds}{d\eta} \rightarrow \begin{cases} \frac{2}{\pi} \epsilon e^\mu [\cos(\beta + \eta_0) - \cos(\beta + \eta_0 - \alpha)], & |\sum_-| \rightarrow \infty, \\ -\frac{2}{\pi} \epsilon e^\mu [\cos(\beta + \eta_0) + \cos(\beta + \eta_0 - \alpha)], & |\sum_+| \rightarrow \infty. \end{cases}$$

At a double anomaly, different behavior is found. For example, if  $\delta_- > 0$  while  $ka$  is fixed and  $\alpha$  varied, so  $\delta_+ + \delta_- = 0$ , it may be shown that

$$(67) \quad \chi^{(1)}(\alpha) \frac{ds}{d\eta} \rightarrow \frac{2iee^\mu}{\pi} [\cos(\beta + \eta_0) + i \cos(\beta + \eta_0 - \alpha)],$$

while, if  $\delta_- > 0$ ,  $\alpha$  is fixed and  $ka$  varied, so  $\delta_- = \delta_+ \tan^2 \frac{1}{2}\alpha$  and  $\tan^2 \frac{1}{2}\alpha = n/m$  then

$$(68) \quad \chi^{(1)}(\alpha) \frac{ds}{d\eta} \rightarrow \frac{2iee^\mu}{\pi} \left[ \left( \frac{m^{\frac{1}{2}} - n^{\frac{1}{2}}}{m^{\frac{1}{2}} + n^{\frac{1}{2}}} \right) \cos(\beta + \eta_0) - \cos(\beta + \eta_0 - \alpha) \right].$$

Further approximations to  $\chi$  could be found in the above manner. The important point, however, is that, whereas in (†) one of the driving terms in the integral equation which determines the second-order approximation to the normal derivative involves a series which may diverge, in the present formulation the corresponding infinite series has terms which tend to zero as  $|\nu|^{-3/2}$ . Thus equation (65) is a valid approximation if  $\epsilon$  is sufficiently small, regardless of whether or not the conditions (5) obtain.

### 3.3.2. Approximation to the Far-field Amplitudes

An approximation to the amplitudes  $a_\nu^\pm$  may be found from equations (13) and (26). Then  $a_\nu^{+(0)} = a_\nu^{-(0)} \equiv a_\nu^{(0)}$ , say, where

$$(69) \quad a_\nu^{(0)} = -\frac{1}{2}i\sigma^{(0)} / (\kappa a),$$

with  $\sigma^{(0)}$  defined by equation (56).

If  $\kappa_n \rightarrow 0$ ,  $n > 0$ , then  $|\sum_-| \rightarrow \infty$  and  $\kappa_n a \sum_- \rightarrow 2$ . Thus, in the limit

$$(70) \quad a_n^{(0)} \rightarrow -\lim[\xi_0(\alpha) / \xi_0(0)] = -1$$

(since  $ka \cos \alpha = ka - 2n\pi$ ), while all other  $a_\nu^{(0)}$  vanish.

Also if  $\kappa_{-m} \rightarrow 0$ ,  $m > 0$ , and  $|\Sigma_+| \rightarrow \infty$ , then

$$(71) \quad a_{-m}^{(0)} \rightarrow -\lim[\xi_0(\alpha)/\xi_0(\pi)] = -1,$$

all other  $a_\nu^{(0)}$  vanishing in the limit.

At a double anomaly,  $\kappa_n$  and  $\kappa_{-m} \rightarrow 0$ ,  $|\Sigma_\pm| \rightarrow \infty$ , and  $\xi_0(\alpha) = \xi_0(0) = \xi_0(\pi)$ , because  $ka \cos \alpha$  is a multiple of  $\pi$ . Then

$$(72) \quad a_n^{(0)} \rightarrow -1/(1+\tau),$$

where

$$(73) \quad \tau = \frac{1}{2} \lim \kappa_n a \Sigma_+.$$

$\tau$  thus depends on the manner in which the limit is approached (see the paragraph preceding equation (42)). For if  $ka$  is fixed, and  $\alpha$  varied, then  $\tau = \mp i$  accordingly as  $\kappa_n \rightarrow 0$  through real or imaginary values respectively, while if  $\alpha$  is fixed and  $ka$  varied, then  $\tau = \tan \frac{1}{2}\alpha$ . Therefore, in the former case there is an abrupt phase shift in  $a_n^{(0)}$  of  $\pi/2$  as the double anomaly is passed. Also

$$(74) \quad a_{-m}^{(0)} \rightarrow -\tau/(1+\tau).$$

It appears then, for a grating of small elements on which is satisfied the boundary condition  $u = 0$ , that near a Rayleigh wavelength the scattered field consists almost entirely of one, or at most two nonevanescent waves which propagate along the surface of the grating.

Additional approximations to  $a_\nu^{(\pm)}$  may be found, but  $a_\nu^{(0)}$  exhibits the chief interaction effects. In fact, equation (64) implies that the next approximation,  $a_\nu^{(1)}$ , is equal to  $a_\nu^{(0)}$ .

An explicit expression for  $a_\nu^{(0)}$  is readily found, with the help of equations (53), (54), and (56):

$$(75) \quad a_\nu^{(0)} = -\frac{2}{\kappa_n a} \frac{2i(p+\mu)/\pi + \Sigma_0}{[2i(p+\mu)/\pi + \Sigma_0(\alpha, ka)][2i(p+\mu)/\pi + \Sigma_0 + \Sigma_+ + \Sigma_-]},$$

in which

$$(76) \quad \Sigma_0 \equiv \Sigma_0(0, ka) = \Sigma_0(\pi, ka).$$

The behavior of  $a_n^{(0)}$  near  $\kappa_n = 0$  is determined essentially by the quantity  $(\kappa_n a)^{-1}[2i(p+\mu)/\pi + \Sigma_0 + \Sigma_+ + \Sigma_-]^{-1}$ . If it is assumed that  $|\Sigma_-| \rightarrow \infty$  (so  $\delta_- \rightarrow 0$ ) and  $\Sigma_0 + \Sigma_+ = r + is$ , then, since  $\Sigma_- \doteq [2/(ka\delta_-)]^{\frac{1}{2}}$ , it is not difficult to show that

$$(77) \quad \frac{d}{d\delta_-} |a_n^{(0)}|^2 = \begin{cases} -\infty \operatorname{sgn}(1+r), & \delta_- = 0+, \\ -\infty \operatorname{sgn}\left[s + \frac{2}{\pi} \log\left(\frac{1}{4} \gamma e^\mu\right)\right], & \delta_- = 0-. \end{cases}$$

Since  $|r|$ ,  $|s| \ll 1$  if  $ka$  is sufficiently large, while the logarithm is negative when  $\epsilon e^\mu$  is small,  $|a_n^{(0)}|^2$  will, under these conditions, exhibit a cusp and a maximum at a Rayleigh wavelength. With decreasing  $ka$  this behavior may possibly change.

If the derivatives of  $|a_\nu^{(0)}|^2$ , ( $\nu \neq n$ ) are calculated, ( $a_\nu^{(0)} \rightarrow 0$ ), it is found that

$$(78) \quad \frac{d}{d\delta_-} |a_\nu^{(0)}|^2 = \pm \frac{2ka}{|\kappa_\nu a|^2}$$

accordingly as  $\delta_- = 0 \pm$ . Thus  $|a_{\nu}^{(0)}|^2$ , ( $\nu \neq n$ ), exhibits a cusp with finite slope independent of  $n$ , at a Rayleigh wavelength.

Near a double anomaly, ( $\kappa_n, \kappa_{-m} \rightarrow 0$ ), for  $ka$  fixed (but not necessarily large) and  $\alpha$  variable, then the derivatives of  $|a_n^{(0)}|^2$ , ( $n > 0$ ), with respect to  $\delta_-$  have opposite signs as  $\delta_- \rightarrow 0 \pm$ . In general,  $|a_n^{(0)}|^2$  possesses a cusp, with infinite slope. Also, for  $\nu \neq n, -m$ ,

$$(79) \quad \frac{d}{d\delta_-} |a_{\nu}^{(0)}|^2 = \pm \frac{ka}{|\kappa_{\nu} a|^2},$$

accordingly as  $\delta_- = 0 \pm$ .

Moreover, in the neighborhood of a double anomaly, with  $\alpha$  fixed and  $ka$  varied, the derivatives of  $|a_n^{(0)}|^2$  are given by equation (77), where now  $\Sigma_0 = r + is$ . The slopes of the amplitudes which vanish at the anomaly satisfy

$$(80) \quad \frac{d}{d\delta_-} |a_{\nu}^{(0)}|^2 = \pm \frac{2ka}{|\kappa_{\nu} a|^2 (1 + \tan \frac{1}{2}\alpha)^2},$$

accordingly as  $\delta_- = 0 \pm$ .

In general, then (and certainly for sufficiently large values of  $ka$ ), the intensities of all the plane waves exhibit cusps, with finite or infinite slope, at Rayleigh wavelengths. Only the wave (or waves) propagated tangentially along the grating possesses an intensity maximum at a Rayleigh wavelength; the amplitudes of all other scattered plane waves fall to zero at these points.

A particular example of the intensity variations near an anomaly is illustrated in Fig. 4. It is assumed that  $\epsilon e^{\mu} = 0.01$ ;  $ka \div 5\pi/2$  and  $\cos \alpha \div 1/5$ , so  $ka(1 - \cos \alpha) = 2\pi + \delta_-$ ,  $|\delta_-| \ll 1$ . Thus the chosen anomaly corresponds to  $n = 1$ .

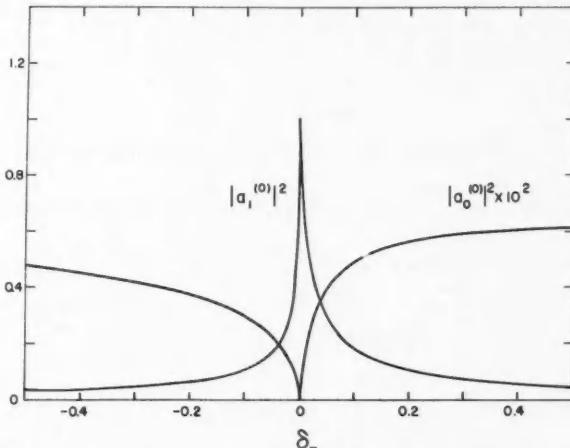


FIG. 4. Variation near a Rayleigh wavelength ( $n = 1$ ) of  $|a_0^{(0)}|^2$  and  $|a_1^{(0)}|^2$  for a grating of small, elliptic cylinders. The derivative of  $|a_0^{(0)}|^2$  with respect to  $\delta_-$  is finite, although discontinuous, at  $\delta_- = 0$ ; the corresponding derivative of  $|a_1^{(0)}|^2$  is infinite.

The intensities  $|a_0^{(0)}|^2$  and  $|a_1^{(0)}|^2$  are exhibited as functions of  $\delta_-$  near  $\delta_- = 0$ .  $|a_1^{(0)}|^2$  tends to unity,  $|a_0^{(0)}|^2$  to zero as  $\delta_-$  approaches zero. Both possess cusps at  $\delta_- = 0$ . The calculations leading to the curves of Fig. 4 are based on equation (75), together with certain simplifying assumptions, namely: that  $\Sigma_0$  and  $\Sigma_0(\alpha, ka)$  may be neglected; that  $\Sigma_-$  is determined by equation (37), while  $\Sigma_+$  is constant and equal to  $-0.122 (1-i)$ , its value for  $\delta_- = 0$ . For small  $|\delta_-|$  and the selected  $ka$ , these assumptions appear plausible.

It may be seen that  $|a_1^{(0)}|^2$  is greater than 0.5 only in a band of width 0.02 in  $\delta_-$ . In the present circumstances, this corresponds to a variation in  $\alpha$  of about  $0.2^\circ$ , or a fractional change in wavelength of approximately 0.003.

#### REFERENCES

ARTMANN, K. 1942. Z. Physik, **119**, 529.  
 JAHNKE, E. and EMDE, F. 1945. Tables of functions with formulae and curves, 4th ed. (Dover Publications, New York).  
 JEFFREYS, H. and JEFFREYS, B. S. 1956. Methods of mathematical physics, 3rd ed. (Cambridge University Press).  
 KARP, S. N. 1955. Diffraction by an infinite grating of arbitrary cylinders. New York University, Inst. of Mathematical Sciences, Division of Electromagnetic Research, Research Report No. EM-85.  
 KARP, S. N. and RADLOW, J. 1956. IRE Trans. **AP-4**, 654.  
 LIPPmann, B. A. 1953. J. Opt. Soc. Am. **43**, 408.  
 MEECHAM, W. C. 1956. J. Appl. Phys. **27**, 361.  
 MILLAR, R. F. 1960. Can. J. Phys. **38**, 272.  
 PALMER, C. H. 1952. J. Opt. Soc. Am. **42**, 269.  
 RAYLEIGH, LORD. 1907. Proc. Roy. Soc. A, **79**, 399.  
 TITCHMARSH, E. C. 1948. Introduction to the theory of Fourier integrals, 2nd ed. (Oxford University Press).  
 TWERSKY, V. 1952. J. Appl. Phys. **23**, 1099.  
 ——— 1956. IRE Trans. **AP-4**, 330.  
 ——— 1957. Notes on scattering by gratings. Sylvania Electronic Defense Laboratory, Mountain View, California. Technical Memorandum EDL-M105.  
 ——— 1958. On scattering of waves by the infinite grating of circular cylinders. Sylvania Electronic Defence Laboratory, Mountain View, California, Engineering Report EDL-E28.  
 ——— 1960. On multiple scattering of waves. Sylvania Electronic Defense Laboratories, Mountain View, California, No. EDL-L16.  
 WAIT, J. R. 1955. Appl. Sci. Research, B, **4**, 393.  
 WOOD, R. W. 1902. Phil. Mag. Ser. 6, **4**, 396.  
 ——— 1912. Phil. Mag. Ser. 6, **23**, 310.  
 ——— 1935. Phys. Rev. **48**, 928.

#### APPENDIX

##### A MORE RAPIDLY CONVERGENT EXPANSION FOR $\Sigma_{\pm}$

The functions  $\Sigma_{\pm}$  are defined by equation (18). Consider the sum

$$\sum_{n=1}^{\infty} \frac{e^{in\delta}}{n^4}$$

where, for the present,  $0 < \delta < 2\pi$ .

Let

$$f(x) = \begin{cases} e^{ix^{\frac{1}{2}}}/x^{\frac{1}{2}}, & x \geq 1, \\ 0, & x < 1. \end{cases}$$

Then

$$F_c(y) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty f(x) \cos xy \, dx, \quad y > 0,$$

$$= \begin{cases} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left( \frac{F[\sqrt{(y+\delta)}]}{\sqrt{y+\delta}} + \frac{F[\sqrt{(\delta-y)}]}{\sqrt{\delta-y}} \right), & \delta > y, \\ \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left( \frac{F[\sqrt{(y+\delta)}]}{\sqrt{y+\delta}} + \frac{F^*[\sqrt{(y-\delta)}]}{\sqrt{y-\delta}} \right), & 0 < \delta < y, \end{cases}$$

where  $F(z) = \int_z^\infty e^{\alpha \lambda^2} d\lambda$ , and  $(*)$  denotes complex conjugate.  $f(x)$  satisfies the conditions of Titchmarsh (1948, Theorem 47, p. 67). Therefore, with  $\alpha = 1$ ,  $\beta = 2\pi$ ,

$$\frac{1}{2} \sum_{n=1}^{\infty} [f(n+0) + f(n-0)] =$$

$$2 \left\{ \frac{F(\sqrt{\delta})}{\sqrt{\delta}} + \sum_{n=1}^{\infty} \left( \frac{F[\sqrt{(2\pi n + \delta)}]}{\sqrt{(2\pi n + \delta)}} + \frac{F^*[\sqrt{(2\pi n - \delta)}]}{\sqrt{(2\pi n - \delta)}} \right) \right\}.$$

Thus

$$\sum_{n=1}^{\infty} \frac{e^{in\delta}}{n^{\frac{1}{2}}} = \frac{1}{2} e^{i\delta} + 2 \left\{ \frac{F(\sqrt{\delta})}{\sqrt{\delta}} + \sum_{n=1}^{\infty} \left( \frac{F[\sqrt{(2\pi n + \delta)}]}{\sqrt{(2\pi n + \delta)}} + \frac{F^*[\sqrt{(2\pi n - \delta)}]}{\sqrt{(2\pi n - \delta)}} \right) \right\},$$

or

$$\sum_{n=1}^{\infty} \frac{e^{in\delta}}{n^{\frac{1}{2}}} = \frac{1}{2} e^{i\delta} + 2 \left\{ \sum_{-\infty < n < \frac{\delta}{2\pi}} \frac{F[\sqrt{(\delta - 2\pi n)}]}{\sqrt{(\delta - 2\pi n)}} + \sum_{\frac{\delta}{2\pi} < n < \infty} \frac{F^*[\sqrt{(2\pi n - \delta)}]}{\sqrt{(2\pi n - \delta)}} \right\},$$

from which  $\sum_{\pm}$  are readily found by appropriate choice of  $\delta$ .

In this latter formula,  $\delta$  may take any value. If  $\delta$  is sufficiently small, then

$$\sum_{n=1}^{\infty} \frac{e^{in\delta}}{n^{\frac{1}{2}}} \doteq \begin{cases} 2 \frac{F(\sqrt{\delta})}{\sqrt{\delta}}, & \delta > 0, \\ 2 \frac{F^*(\sqrt{|\delta|})}{\sqrt{|\delta|}}, & \delta < 0, \end{cases}$$

which is the behavior found by Karp and Radlow (1956).

## SCATTERING BY A GRATING. II<sup>1</sup>

R. F. MILLAR

### ABSTRACT

The analysis of Part I is repeated for the second boundary value problem in which the normal derivative of the wave function vanishes on the elements of the grating; the existence of anomalies is again inferred. The behavior of the spectral intensity near an anomaly for a grating of small elliptic cylinders is considered and found to be much different from the case studied in Part I.

The infinite reflection grating is discussed. Weak anomalies are predicted in the spectra for the case in which the wave function vanishes on the grating; strong (Wood) anomalies are found in the spectra for the second problem.

The finite transmission grating is treated very approximately. Anomalous behavior, dependent now on the number of elements, is inferred.

### INTRODUCTION

The study initiated in Part I is continued. In Section 1, a less detailed development is given for the second boundary value problem. The infinite reflection grating is discussed in Section 2, while the finite grating is considered in Section 3.

#### 1. $\partial u / \partial n = 0$ ON THE CYLINDERS

##### 1.1. *Derivation and Formal Solution of the Integral Equations*

The notation of the previous part (henceforth referred to as I) will be employed in the present discussion.

Equation (I,1) gives, for the total field at a point  $P$  not on one of the cylinders,

$$(1) \quad u(P) = u^1(P) + \frac{1}{4}i \sum \int_{C_v} u \frac{\partial}{\partial n_v} H(kr_v) ds_v.$$

If  $P$  is now permitted to tend to a point  $P_\lambda$  on  $C_\lambda$ , then the integral equation satisfied by  $u$  is seen to be

$$(2) \quad 2iu(P_\lambda) + \sum \int_{C_v} u \frac{\partial}{\partial n_v} H(kr_{v\lambda}) ds_v = 4iu^1(P_\lambda),$$

where the integral for  $v = \lambda$  is a principal value.

(It may be noted that, if at the outset the cylinders are assumed to be coplanar strips, then equations (2) reduce to the identities

$$u(P_\lambda^+) + u(P_\lambda^-) = 2u^1(P_\lambda),$$

where  $P_\lambda^\pm$  represent limiting values as  $P \rightarrow P_\lambda$  through values of  $z \gtrless 0$ , respectively. The appropriate analysis leads to a set of differential-integral equations, but this case will not be discussed here.)

<sup>1</sup>Manuscript received September 6, 1960.

Contribution from the Radio and Electrical Engineering Division, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 6106.

The equations (2) may be modified in a manner similar to that employed in I. Then, if  $\phi_r \equiv u(P_r)$ ,

$$(3) \quad \phi_\lambda = 4i\zeta_\lambda(\alpha) - e^{-i\lambda ka} \zeta_\lambda(0) \sum_{\nu < \lambda} c(\nu - \lambda) \tau_\nu(\alpha, 0) \\ - e^{i\lambda ka} \zeta_\lambda(\pi) \sum_{\nu > \lambda} c(\nu - \lambda) \tau_\nu(\alpha, \pi),$$

where  $\zeta_\lambda(\alpha)$  is the solution of

$$(4) \quad 2i\zeta_\lambda(\alpha) + \sum \int_{C_\nu} \zeta_\nu(\alpha) \frac{\partial}{\partial n_\nu} G(\nu, \lambda) ds_\nu = u^1(P_\lambda),$$

$$(5) \quad \tau_\nu(\alpha, \theta) \equiv \int_{C_\nu} \phi_\nu \frac{\partial}{\partial n_\nu} e^{-ik(y_\nu \cos \theta + z_\nu \sin \theta)} ds_\nu,$$

and  $G(\nu, \lambda)$  is defined by equation (I,9). The physical interpretation of the above equations is analogous to that of equations (I,11) and (I,12). The functions  $\tau_\nu(\alpha, \theta)$  are determined by

$$(6) \quad \tau_\lambda(\alpha, \theta) = 4iq_\lambda(\alpha, \theta) - e^{-i\lambda ka} q_\lambda(0, \theta) \sum_{\nu < \lambda} c(\nu - \lambda) \tau_\nu(\alpha, 0) \\ - e^{i\lambda ka} q_\lambda(\pi, \theta) \sum_{\nu > \lambda} c(\nu - \lambda) \tau_\nu(\alpha, \pi),$$

where

$$(7) \quad q_\lambda(\alpha, \theta) \equiv \int_{C_\lambda} \zeta_\lambda(\alpha) \frac{\partial}{\partial n_\lambda} e^{-ik(y_\lambda \cos \theta + z_\lambda \sin \theta)} ds_\lambda.$$

Sets of linear algebraic equations for the determination of  $\tau_\nu(\alpha, 0)$ ,  $\tau_\nu(\alpha, \pi)$  may be found from equations (6) by setting  $\theta$  equal to 0 and  $\pi$  in turn. As was the case in I, the rigorous solution of the set of equations would appear to be practically impossible unless  $N$  is small, or the grating infinite. This latter case will be next considered.

### 1.2. The Infinite Grating of Identical Cylinders

When the grating is infinite, and the cylinders identical, then the number of unknowns in the linear algebraic equations is reduced to two ( $\tau(\alpha, 0)$  and  $\tau(\alpha, \pi)$ , say, where the subscript "0" has been omitted). These may be found from equations (I,17) by the transformation  $\sigma \rightarrow -\tau$ ,  $p \rightarrow q$  (see equations (18) of Part II). Then the field  $[\phi(\alpha)]$  on  $C_0$  is determined by equation (3), with  $\lambda = 0$ :

$$(8) \quad \phi(\alpha) = 4i\zeta(\alpha) - \zeta(0)\tau(\alpha, 0) \sum_- - \zeta(\pi)\tau(\alpha, \pi) \sum_+.$$

The field of the grating may be found directly from equation (I,24):

$$(9) \quad u(P) = u^1(P) + \sum b_\nu^\pm e^{ik(y \cos \theta_\nu \pm z \sin \theta_\nu)},$$

the  $\pm$  sign being chosen accordingly as  $z \gtrless 0$ , it being understood that  $P$  lies outside a strip containing the cylinders. Here

$$(10) \quad b_v^\pm = \frac{1}{2}i\tau(\alpha, \pm\theta_v)/(\kappa a),$$

and the angle  $\theta_v$  is defined by equations (I,27).

Limiting forms of  $b_v^\pm$  at a Rayleigh wavelength may be found by application of the transformation  $a_v^\pm \rightarrow b_v^\pm$ ,  $\sigma \rightarrow -\tau$ ,  $p \rightarrow q$  to equations (I,28) through (I,32).

A comparison of equations (I,26) and (10) makes evident the possibility of similarity in the behavior of  $a_v^\pm$  and  $b_v^\pm$  near a Rayleigh wavelength. For example, Wood (1902, 1935) observed bright and dark narrow lines, as well as intensity variations somewhat like Fig. 2(d) of I in the  $H$ -polarized spectra of a reflection grating. (In Section 2, it will be seen that, for this polarization ( $\partial u / \partial n = 0$  on the cylinders), the spectra of reflection and transmission gratings are qualitatively similar.) One concludes, therefore, that an infinite transmission grating with identical elements of arbitrary form upon which is satisfied either of the boundary conditions  $u = 0$  or  $\partial u / \partial n = 0$  may be expected to exhibit anomalous behavior near the Rayleigh wavelengths. The precise form of the anomalies is, of course, dependent upon the shape, size, and spacing of the elements, and the boundary condition, but the discussion, which follows equation (I,32), of the possible variation of  $a_v^\pm$  near a Rayleigh wavelength is equally valid for the coefficients  $b_v^\pm$ .

The equation satisfied by  $\xi(\alpha)$  may be obtained from equation (4):

$$(11) \quad 2i\xi + \int_{C_0} ds\xi \frac{\partial}{\partial n} \sum G_v e^{ivka \cos \alpha} = u^1(P_0),$$

where  $G_v$  is defined by equations (I,45), the normal derivative operating on  $y, z$ . For the case of small, elliptic cylinders (see Fig. 1 of I) it is assumed that

$$(12) \quad \xi(\alpha) = \sum_{n=0}^{\infty} \xi_n(\alpha) \epsilon^n,$$

where again  $\epsilon = kb \ll 1$ , and  $2b$  is the interfocal distance. Then  $\xi_0, \xi_1, \dots$ , satisfy the equations

$$(13) \quad \left\{ \begin{array}{l} 2i\xi_0 + \int_{C_0} ds\xi_0 \frac{\partial}{\partial n} g_0 = u_0^1(P_0), \\ 2i\xi_1 + \int_{C_0} ds\xi_1 \frac{\partial}{\partial n} g_0 = u_1^1(P_0) - \int_{C_0} ds\xi_0 \frac{\partial}{\partial n} g_1, \\ \vdots \\ \vdots, \end{array} \right.$$

in which the integrals are principal values, while the functions  $g_n$  are defined by the third of equations (I,49).

The normal derivative of  $g_0$  which appears in (13) may be found from equations (I,51) and (34†), where (†) again refers to Millar (1960):

$$(14) \quad ds \frac{\partial}{\partial n} g_0 = \frac{2i}{\pi} d\eta \frac{\partial}{\partial \mu} \psi_0 = \frac{2i}{\pi} d\eta \left[ \frac{1}{2} + \sum_{n=1}^{\infty} e^{-2n\mu} \cos n(\eta + \eta_0) \right].$$

Because  $u_0^1(P_0) = 1$ , it follows that

$$(15) \quad \xi_0(\alpha) = -\frac{1}{4}i.$$

$\xi_1(\alpha)$  may be determined in like manner. According to equations (I,59),

$$(16) \quad ds \frac{\partial}{\partial n} g_1 = -\frac{\partial Y}{\partial \mu} d\eta \sum_1 (\alpha, ka),$$

which implies the vanishing of the integral on the right-hand side of the second of equations (13). Then

$$(17) \quad \xi_1(\alpha) = \frac{1}{4}e^{\mu} \cos(\eta - \alpha + \beta).$$

It is apparent that, to terms of order  $\epsilon$  inclusive, the function  $\xi(\alpha)$  does not differ from the corresponding solution for the field on an isolated cylinder.

To determine an approximation to  $\phi(\alpha)$  (equation (8)), it is necessary to calculate  $\tau(\alpha, 0)$  and  $\tau(\alpha, \pi)$  to the same order. As mentioned previously, these functions are obtained from equations (I,17) by the transformation  $\sigma \rightarrow -\tau$ ,  $p \rightarrow q$ , with  $q$  defined by equation (7):

$$(18) \quad \begin{cases} \tau(\alpha, 0) = -4i \frac{q(\alpha, \pi)q(\pi, 0)\sum_+ - q(\alpha, 0)[1 + q(\pi, \pi)\sum_+]}{[1 + q(0, 0)\sum_-][1 + q(\pi, \pi)\sum_+] - q(\pi, 0)q(0, \pi)\sum_+\sum_-} \\ \tau(\alpha, \pi) = -4i \frac{q(\alpha, 0)q(0, \pi)\sum_- - q(\alpha, \pi)[1 + q(0, 0)\sum_-]}{[1 + q(0, 0)\sum_-][1 + q(\pi, \pi)\sum_+] - q(\pi, 0)q(0, \pi)\sum_+\sum_-}. \end{cases}$$

It may be shown that the series expansion of  $q(\alpha, \theta)$  begins with the term  $\epsilon^2 q_2(\alpha, \theta)$ , where

$$(19) \quad q_2(\alpha, \theta) = -i \int_0^{2\pi} \xi_1 \frac{\partial}{\partial \mu} (Y \cos \theta + Z \sin \theta) d\eta$$

$$- \frac{1}{2} \int_0^{2\pi} \xi_0 \frac{\partial}{\partial \mu} (Y \cos \theta + Z \sin \theta)^2 d\eta$$

$$= iQ(\alpha, \theta), \text{ say.}$$

$Q(\alpha, \theta)$  is a real function, and  $Y, Z$  are defined by equations (I,48).

At first sight, it would appear that the zero-order approximations  $\tau^{(0)}(\alpha, 0)$  and  $\tau^{(0)}(\alpha, \pi)$  are zero since the  $q$ -functions are of order  $\epsilon^2$ . This is indeed so if both  $\sum_{\pm}$  are negligible, but if, for example,  $|\sum_- \epsilon^2 q_2(0, 0)| \gg 1$  (a condition which may well be realized if  $\epsilon \neq 0$ ), then

$$\tau^{(0)}(\alpha, 0) \sum_- \rightarrow 4i q_2(\alpha, 0) / q_2(0, 0).$$

This quantity, which appears in equation (8), is of order zero in  $\epsilon$ . Thus it is necessary, even in a first (that is, zero-order) approximation, to retain the terms of order  $\epsilon^2$  in numerators and denominators of the expressions (18).  $\phi^{(0)}(\alpha)$  may then be defined (see equations (8), (15), (17)) by

$$(20) \quad \phi^{(0)}(\alpha) = 1 + \frac{1}{4}i[\tau^{(0)}(\alpha, 0) \sum_{-} + \tau^{(0)}(\alpha, \pi) \sum_{+}] + \epsilon e^{\mu} [i \cos(\eta - \alpha + \beta) - \frac{1}{4}[\tau^{(0)}(\alpha, 0) \sum_{-} - \tau^{(0)}(\alpha, \pi) \sum_{+}] \cos(\eta + \beta)].$$

It should be pointed out that, although equations (18) (with  $\tau$ ,  $q$ , replaced by  $\tau^{(0)}$ ,  $\epsilon^2 q_2$ , respectively) are correct to order  $\epsilon^2$  inclusive,  $\phi^{(0)}(\alpha)$  is correct only to order  $\epsilon$ , because the second-order term in  $\zeta(\alpha)$  has been omitted. Nevertheless, sufficient terms have been retained in equation (20) to make equations (5), (20), and the zero-order form of equations (18) consistent in the quantities  $\phi^{(0)}$  and  $\tau^{(0)}$ .

Various special cases arise, dependent upon the magnitudes of  $\sum_{\pm}$ , but these will not be considered.

Because the major effects of interaction are described by  $\phi^{(0)}$ , it is unnecessary to consider further terms.  $\phi^{(0)}$  will be a valid approximation if  $\epsilon$  is sufficiently small regardless of whether or not interaction is of importance.

The first approximation to  $b_{\nu}^{\pm(0)}$  (equation (10)) is

$$(21) \quad b_{\nu}^{\pm(0)} = \frac{1}{2}i\tau^{(0)}(\alpha, \pm\theta_{\nu})/(\kappa_{\nu}a).$$

If  $\kappa_n \rightarrow 0$ ,  $n > 0$ , then  $\theta_n \rightarrow 0$ , and

$$(22) \quad b_n^{\pm(0)} = \frac{1}{2}i \left[ \frac{\tau^{(0)}(\alpha, 0)}{\kappa_n a} \pm \frac{1}{ka} \frac{\partial}{\partial \theta} \tau^{(0)}(\alpha, \theta) \Big|_{\theta=0} + O(\theta_n) \right].$$

Here the relationship  $\kappa_n = k \sin \theta_n$  has been employed.

If  $\kappa_{-m} \rightarrow 0$ ,  $m > 0$ , then  $\theta_{-m} \rightarrow \pi$ , and

$$(23) \quad b_{-m}^{\pm(0)} = \frac{1}{2}i \left[ \frac{\tau^{(0)}(\alpha, \pi)}{\kappa_{-m} a} \mp \frac{1}{ka} \frac{\partial}{\partial \theta} \tau^{(0)}(\alpha, \theta) \Big|_{\theta=\pi} + O(\pi - \theta_{-m}) \right].$$

The behavior of  $b_{\nu}^{\pm(0)}$  when  $\nu \neq n, -m$ , may be found by consideration of equations (21) and (6):

$$(24) \quad b_{\nu}^{\pm(0)} = \frac{1}{2}i(\kappa_{\nu}a)^{-1}\epsilon^2[4iq_2(\alpha, \pm\theta_{\nu}) - q_2(0, \pm\theta_{\nu})\tau^{(0)}(\alpha, 0)\sum_{-} - q_2(\pi, \pm\theta_{\nu})\tau^{(0)}(\alpha, \pi)\sum_{+}].$$

Equations (24) imply that the  $b_{\nu}^{\pm(0)}$  are, in general (but not everywhere, as will be shown subsequently), of the second order in  $\epsilon$ . It has previously been indicated (see the discussion preceding equation (20)) that, in the limit,  $b_n^{\pm(0)}$  and  $b_{-m}^{\pm(0)}$  are of order zero in  $\epsilon$ .

The expressions for  $b_{\nu}^{\pm(0)}$  are rather more complex than the coefficients  $a_{\nu}^{\pm(0)}$ . Nonetheless, a few remarks may be made about their behavior near, say,  $\kappa_n = 0$ , where  $|\sum_{-}| \gg 1$ ; for according to equations (6) and (18),

$$(25) \quad -\frac{1}{4}i\tau^{(0)}(\alpha, \theta) = \epsilon^2[iQ(\alpha, \theta) - \epsilon^2([Q(\alpha, \theta)Q(\pi, \pi) - Q(\alpha, \pi)Q(\pi, \theta)]\sum_{+} + [Q(\alpha, \theta)Q(0, 0) - Q(\alpha, 0)Q(0, \theta)]\sum_{-}) + \{1 + i\epsilon^2[Q(\pi, \pi)\sum_{+} + Q(0, 0)\sum_{-}]\} + O(\epsilon^6),$$

where  $Q(\alpha, \theta)$  is the (real) function defined by equation (19).

If  $|\Sigma_-| \gg 1$ , while  $|\Sigma_+|$  is small by comparison, then

$$(26) \quad -\frac{1}{4}i\tau^{(0)}(\alpha, \theta) \doteq \frac{i\epsilon^2 Q(\alpha, \theta) + \epsilon^4 [Q(\alpha, \theta)Q + Q(\alpha, 0)Q(0, \theta)]\Sigma_-}{1 - i\epsilon^2 Q(\Sigma_+ + \Sigma_-)},$$

where  $Q > 0$  is defined by

$$(27) \quad Q \equiv -Q(0, 0) = -Q(\pi, \pi) = \frac{1}{4}\pi(\cosh 2\mu - \cos 2\beta).$$

The denominator of the expression (26) may become quite small. For if  $\delta_- < 0$  (where  $\delta_- = ka(1 - \cos \alpha) - 2n\pi$ ), and  $|\delta_-|$  is sufficiently small, then  $\Sigma_- \doteq -i[2/(ka|\delta_-|)]^{\frac{1}{2}}$ , and the real part of the denominator vanishes if  $\delta_- = \delta_{\max} < 0$ , where

$$(28) \quad (ka|\delta_{\max}|)^{\frac{1}{2}} = \sqrt{2\epsilon^2 Q/[1 + \epsilon^2 \operatorname{Im} \Sigma_+]}.$$

Since the imaginary part of its denominator is small, and slowly varying,  $|b_{\nu}^{\pm(0)}|^2$  will exhibit a maximum for  $\delta_- = \delta_{\max}$  determined by equation (28). Then

$$(29) \quad |b_{\nu \max}^{\pm(0)}|^2 = \frac{4}{|\kappa \alpha|^2} \cdot \frac{Q^2(\alpha, 0)Q^2(0, \pm\theta_r)}{Q^4|\Sigma_+|^2} + O(\epsilon^2), \quad \nu \neq n.$$

It is evident that the position of the maximum depends explicitly on neither  $\nu$  nor  $n$ , while the magnitude at the maximum is essentially independent of  $\epsilon$ .

The numerator of  $\tau^{(0)}(\alpha, \pm\theta_r)$  may likewise be examined, and there arises the possibility of minima in both  $|b_{\nu}^{+(0)}|^2$  and  $|b_{\nu}^{-(0)}|^2$ . For  $\tau^{(0)}(\alpha, +\theta_r)$  may be approximately zero when  $\delta_- = \delta_{\min}^+ < 0$ , where

$$(30) \quad (ka|\delta_{\min}^+|)^{\frac{1}{2}} = \sqrt{2\epsilon^2 [Q + Q(\alpha, 0)Q(0, \theta_r)/Q(\alpha, \theta_r)]},$$

the first term of which is essentially  $(ka|\delta_{\max}|)^{\frac{1}{2}}$ , and  $|\delta_{\min}^+|$  is assumed sufficiently small for the above approximation to  $\Sigma_-$  to be valid. For the existence of  $\delta_{\min}^+$ , it is necessary, then, for the right-hand side of this equation to be positive, with  $|Q(\alpha, \theta_r)|$  sufficiently greater than zero. Evidently  $\delta_{\min}^+$  is not independent of  $\nu$ .

The integrations in equations (19) may be performed, and it is seen that

$$(31) \quad Q(\alpha, \theta) = \frac{1}{4}\pi[\sinh 2\mu - e^{2\mu} \cos(\alpha - \theta) + \cos(\alpha + \theta - 2\beta)].$$

Thus  $\delta_{\min}^-$ , which is obtained by changing the sign of  $\theta$ , in the right-hand side of equation (30), is not equal to  $\delta_{\min}^+$ , and the positions of minimum intensity of corresponding waves in  $z \geq 0$  are not coincident.

The condition that  $|Q(\alpha, \theta_r)|$  be sufficiently greater than zero is not of necessity satisfied. It may be shown that  $Q(\alpha, \theta) = 0$  if

$$(32) \quad \cos(\chi - \chi_0) = [e^{4\mu} - 2e^{2\mu} \cos 2(\alpha - \beta) + 1]^{-\frac{1}{2}} \sinh 2\mu,$$

where

$$(33) \quad \chi = \theta - \beta,$$

$$(34) \quad \begin{pmatrix} \sin \\ \cos \end{pmatrix} \chi_0 = [e^{4\mu} - 2e^{2\mu} \cos 2(\alpha - \beta) + 1]^{-\frac{1}{2}} (e^{2\mu} \pm 1) \begin{pmatrix} \sin \\ \cos \end{pmatrix} (\alpha - \beta),$$

and the angles  $\chi$ ,  $\chi_0$  are always real. For example, if the cylinders are circular ( $\mu \rightarrow \infty$ ), then  $Q(\alpha, \theta) = 0$  if  $\cos(\alpha - \theta) = \frac{1}{2}$ .

In addition to the condition on  $Q(\alpha, \theta_\nu)$ , it is evident, according to equation (30), that  $Q(\alpha, 0) Q(0, \theta_\nu) / Q(\alpha, \theta_\nu)$  must be sufficiently large that the positions of minimum and maximum are not so close as to invalidate the arguments employed in their derivation. Furthermore, it is not difficult to show that

$$(35) \quad \frac{d}{d\delta_-} |b_\nu^{\pm(0)}|^2 = \begin{cases} -\infty \operatorname{sgn} [\operatorname{Re} \sum_+ + O(\epsilon^2)], & \delta_- = 0+ \\ -\infty \operatorname{sgn} [Q(\alpha, 0) Q(0, \pm\theta_\nu) / Q(\alpha, \pm\theta_\nu)], & \delta_- = 0-. \end{cases}$$

Then if  $Q(\alpha, 0) Q(0, \pm\theta_\nu) / Q(\alpha, \pm\theta_\nu) < 0$ ,  $\delta_{\min}^{\pm}$  lies between  $\delta_{\max}$  and 0, and the slope of  $|b_\nu^{\pm(0)}|^2$  is positively infinite at  $\delta_- = 0-$ . On the other hand, if  $\delta_{\max}$  is between  $\delta_{\min}^{\pm}$  and 0, then the derivative at 0- is  $-\infty$ . Whether or not there is a cusp at  $\delta_- = 0$  is determined by  $\operatorname{sgn}[\operatorname{Re} \sum_+]$ , if  $\operatorname{Re} \sum_+$  is sufficiently large that the terms  $O(\epsilon^2)$  in equation (35) are negligible.

The amplitude  $b_n^{\pm(0)}$  of the wave which tends to grazing as  $\kappa_n \rightarrow 0$  likewise may be examined. It may be noted that the form of equation (22) indicates the existence of a maximum in  $|b_n^{\pm(0)}|^2$  (of order  $\epsilon^{-4}$ ) at a point determined by equation (28). On the other hand, the condition which determines a minimum for  $|b_n^{\pm(0)}|^2$  (that is, the vanishing of certain second-order terms) turns out to be  $(ka\delta_-/2)^{\frac{1}{2}} = -\frac{1}{2}kaQ(\alpha, 0)/Q'(\alpha, 0)$ , where the prime denotes differentiation with respect to  $\theta_n$ . Thus  $\delta_-$  must be positive, so the minimum found for all other amplitudes on the long wavelength side of the anomaly is absent. Moreover, the right-hand side of this expression is not small in general, nor positive of necessity. Nevertheless, it is satisfied, for example, in the case of small circular cylinders ( $\mu \rightarrow \infty$  in equation (31)) for a value of  $\cos \alpha$  slightly less than  $\frac{1}{2}$ . Under these conditions,  $|b_n^{+(0)}|^2$  possesses a minimum on the short wavelength side of the anomaly, and a similar minimum occurs for  $|b_n^{-(0)}|^2$  when  $\cos \alpha$  is somewhat greater than  $\frac{1}{2}$ .

The variation in intensities near a double anomaly may be considered in like manner. However, because both  $|\sum_{\pm}| \rightarrow \infty$  it is necessary to employ the complete expression for  $r^{(0)}(\alpha, \theta)$  determined from equations (6) and (18), rather than the more approximate equation (25). This aspect of the problem therefore will not be considered further.

To illustrate some of the foregoing results, numerical calculations of the intensity variations near an anomaly have been performed. The cylinders are circular, of radii  $r$ . The anomaly chosen is that for which  $\kappa_1 \rightarrow 0$ ; it is assumed that  $ka \neq 5\pi/2$  and  $\cos \alpha \neq 1/5$ , so  $ka(1 - \cos \alpha) = 2\pi + \delta_-$ ,  $|\delta_-| \ll 1$ .

The behavior of  $|b_0^{+(0)}|^2$  is exhibited in Fig. 1 for  $kr = 0.005$  and  $0.05$ ; these values correspond to  $\epsilon e^\mu = 0.01, 0.1$ , respectively. The calculation is based on equation (26), with  $\sum_-$  approximated by  $[2/(ka\delta_-)]^{\frac{1}{2}}$ , and  $\sum_+$  assumed constant and equal to  $-0.122(1-i)$ , its value for  $\delta_- = 0$ . The abscissae of the maxima (and minima) are, for the two cases, in the ratio  $(0.05/0.005)^4 = 10^4$  (equations (28) and (30) indicate that the abscissae of maxima and minima are in this ratio); other than in the neighborhoods of the maxima and minima, the intensities, as functions of  $10^m \delta_-$  differ only by the same factor

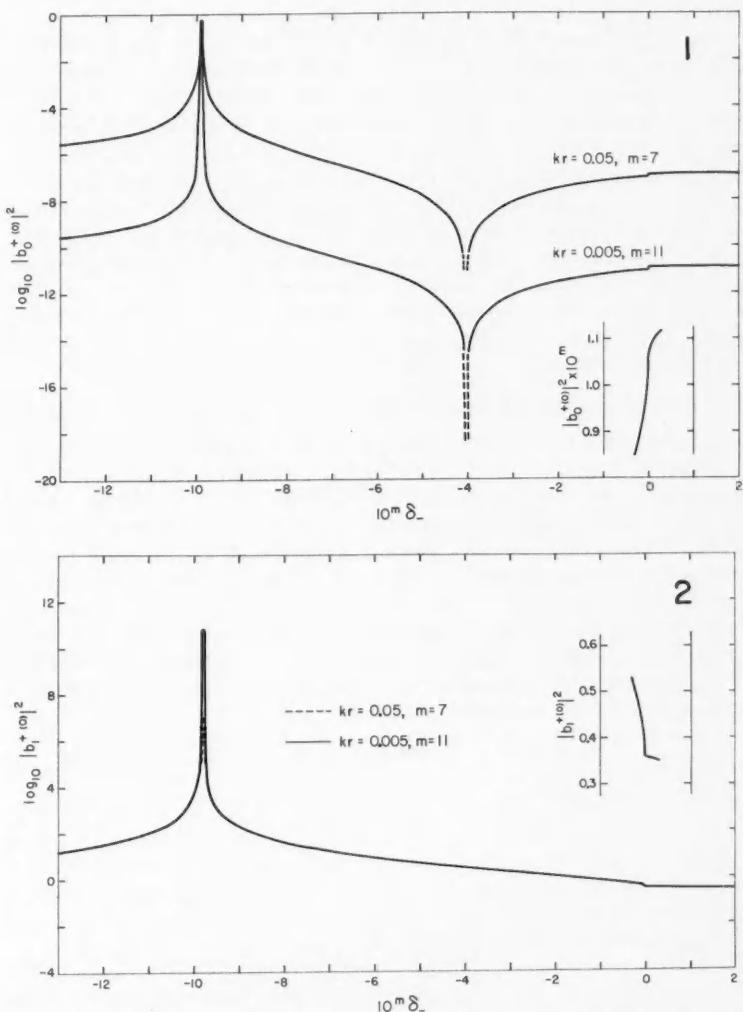


FIG. 1. Variation near a Rayleigh wavelength ( $n = 1$ ) of  $\log_{10}|b_0^{+(0)}|^2$  for gratings of small, circular cylinders of radii  $r$ ;  $kr = 0.005$  and  $0.05$ . The insert illustrates on a linear scale the behavior of  $|b_0^{+(0)}|^2$  near  $\delta_- = 0$ .

FIG. 2. Variation near a Rayleigh wavelength ( $n = 1$ ) of  $\log_{10}|b_1^{+(0)}|^2$  for gratings of small, circular cylinders;  $kr = 0.005$  and  $0.05$ . The insert illustrates on a linear scale the behavior of  $|b_1^{+(0)}|^2$  near  $\delta_- = 0$ .

of  $10^4$ . The magnitudes of the maxima (0.595) are independent of  $kr$  (equation (29)). To the present approximation, the minima are zero; if higher-order terms are considered, the minima will be found to be of order  $(kr)^8$ .

The rapid variation in intensity near  $\delta_- = 0$  is masked by the logarithmic scale in Fig. 1. The insert, drawn to a linear scale, illustrates this phenomenon.

Figure 2 exhibits the behavior of  $|b_1^{(0)}|^2$  (the intensity of the wave which becomes grazing for  $\delta_- = 0$ ), for the same values of the parameters. To carry out the calculation, the expression for  $r^{(0)}(\alpha, \theta_1)$  (equation (26)) was expanded about  $\theta_1 = 0$  to terms of order  $\theta_1$ , inclusive,  $\alpha$  being assumed fixed. For small  $\theta_1$ ,  $\kappa_1 \approx k\theta_1$ , so  $\theta_1 \approx [2\delta_-/(ka)]^{\frac{1}{2}}$ . Other than in the neighborhood of the strong maximum, previously seen to be of order  $(kr)^{-4}$ , the intensity, as a function of  $10^m\delta_-$ , is independent of  $kr$ . (It may be noted that this maximum is not observable, associated as it is with an evanescent wave.) The previous contention that there is no sharp minimum in this case is borne out by the figure. The rapid variation of intensity near  $\delta_- = 0$  is plotted to a linear scale in the insert to Fig. 2.

## 2. THE INFINITE REFLECTION GRATING

As is well known, the results obtained both in I and the present paper may be employed to analyze the behavior of certain reflection gratings.

Suppose that the wave function which vanishes on the cylinders is denoted by  $u_1(\alpha; y, z)$ , while  $u_2(\alpha; y, z)$  represents the wave function whose normal derivative is null on the cylinders. Further, assume that the scatterers forming the grating are symmetrical about the plane  $z = 0$  (so  $\beta = 0$  or  $\frac{1}{2}\pi$  for elliptical cylinders).

Consider the reflection grating defined by the strips ( $S$ ) of the plane  $z = 0$  lying between cylinders, and that portion ( $C$ ) of the cylinders lying in  $z \leq 0$ . Then the solutions (defined for  $z \leq 0$ ) to two infinite reflection grating problems are determined by the functions

$$(36) \quad \begin{cases} U_1(\alpha; y, z) = u_1(\alpha; y, z) - u_1(-\alpha; y, z), \\ U_2(\alpha; y, z) = u_2(\alpha; y, z) + u_2(-\alpha; y, z), \end{cases}$$

where  $U_1$ ,  $U_2$  satisfy the following boundary conditions:

- (i)  $U_1 = 0$  on  $C+S$
- (ii)  $\partial U_2 / \partial n = 0$  on  $C+S$ .

Cases (i) and (ii) correspond respectively to scattering of  $E$ - and  $H$ -polarized waves by a perfectly conducting reflection grating.

Through considerations of symmetry, together with equations (I,10), (I,14), (I,26), and (I,44), the following relationships may be proved for a grating with elements symmetrical about the plane  $z = 0$ :

$$(37) \quad \begin{cases} \chi(\alpha; y, z) = \chi(-\alpha; y, -z), \\ \xi(\alpha; y, z) = \xi(-\alpha; y, -z), \\ \sigma(\alpha, \theta) = \sigma(-\alpha, -\theta), \\ p(\alpha, \theta) = p(-\alpha, -\theta), \\ a_s^+(\alpha) = a_s^-(\alpha). \end{cases}$$

In a similar manner (see equations (5), (7), (10), and (11)), it may be shown that

$$(38) \quad \begin{cases} \phi(\alpha; y, z) = \phi(-\alpha; y, -z), \\ \xi(\alpha; y, z) = \xi(-\alpha; y, -z), \\ \tau(\alpha, \theta) = \tau(-\alpha, -\theta), \\ q(\alpha, \theta) = q(-\alpha, -\theta), \\ b_\nu^+(\alpha) = b_\nu^-(\alpha). \end{cases}$$

The above relationships are useful in the discussion of cases (i) and (ii). For simplicity, only the amplitudes of the plane wave spectra in the region  $z < 0$  below the protuberances will be considered.

(i) The field below the protuberance is, according to equation (I,25)

$$(39) \quad U_1(\alpha; y, z) = e^{ik(y \cos \alpha + z \sin \alpha)} - [1 - A_0(\alpha)] e^{ik(y \cos \alpha - z \sin \alpha)} + \sum_{\nu \neq 0} A_\nu(\alpha) e^{ik(y \cos \theta_\nu - z \sin \theta_\nu)},$$

where the equality of  $\theta_0$  and  $\alpha$  (equations (I,27)) has been invoked, and

$$(40) \quad A_\nu(\alpha) = a_\nu^-(\alpha) - a_\nu^-(-\alpha) = 2[p(\alpha, \theta_\nu) - p(\alpha, -\theta_\nu)]/(\kappa a),$$

the latter equality being a consequence of the symmetry relations (37). Thus there is no explicit dependence on  $\sum_\pm$  and much of the anomalous behavior attributable to these functions will be absent from the spectrum of the reflection grating. However, in the neighborhood of a Rayleigh wavelength, the function  $p(\alpha, \theta)$  does depart to some extent from the polar diagram of an isolated cylinder; thus weaker anomalies might be expected, and such have been observed by Wood (1912, 1935) and Palmer (1952). Theoretical work of Twersky (1956) and Meecham (1956) also indicates minor anomalies.

The limiting form of  $A_\nu(\alpha)$  when  $\kappa a \rightarrow 0$  may be found from equation (40). In particular, for  $\nu = n > 0$ ,

$$(41) \quad A_n(\alpha) \rightarrow 4/(ka) \lim p'(\alpha, 0),$$

where the prime indicates differentiation with respect to  $\theta_n$ .

If the protuberances are small and elliptical, then equations (I,55) and (I,63) imply that  $A_\nu(\alpha) = O(\epsilon^2)$ , and in equation (39) only the incident and specularly reflected terms a. z appreciable.

(ii) In this second case, the field is of the form

$$(42) \quad U_2(\alpha; y, z) = e^{ik(y \cos \alpha + z \sin \alpha)} + [1 + B_0(\alpha)] e^{ik(y \cos \alpha - z \sin \alpha)} + \sum_{\nu \neq 0} B_\nu(\alpha) e^{ik(y \cos \theta_\nu - z \sin \theta_\nu)},$$

where

$$(43) \quad B_\nu(\alpha) = b_\nu^-(\alpha) + b_\nu^-(-\alpha) = -[2[q(\alpha, \theta_\nu) + q(\alpha, -\theta_\nu)] + iq(0, -\theta_\nu) \tau(\alpha, 0) \sum_- + iq(\pi, -\theta_\nu) \tau(\alpha, \pi) \sum_+] / (\kappa a),$$

the final expression being a consequence of the relations (38), and equations (6) and (10). Thus  $B_r(\alpha)$  will exhibit anomalies similar to those considered in Section 1 of the present paper. In particular, for small elliptic protuberances, there will be a maximum of  $|B_r(\alpha)|^2$  for separation, wavelength, and angle of incidence determined by equation (28). However, the position of the minimum of  $|B_r(\alpha)|^2$  is not given by equation (30). It is readily shown that, for the reflection grating, the value of  $\delta_{\min} (< 0)$  is determined by

$$(44) \quad (ka|\delta_{\min}|)^{\frac{1}{2}} = \sqrt{2\epsilon^2\{Q + 2Q(\alpha, 0)Q(0, \theta_r)/[Q(\alpha, \theta_r) + Q(\alpha, -\theta_r)]\}}.$$

As may be observed from equations (27) and (31),  $\delta_{\max}$  and  $\delta_{\min}$  are both dependent upon the value of  $\beta$  (0 or  $\frac{1}{2}\pi$ ).

The existence of maxima in the intensities has also been demonstrated by Artmann (1942), Karp and Radlow (1956), and Twersky (1956). The position of the minima appears to have been noted only by Twersky (1956).

If the protuberances are semicircular, with radii  $r$  (where  $kr \ll 1$ ), then

$$(45) \quad (ka|\delta_{\max}|)^{\frac{1}{2}} = \frac{\pi}{2\sqrt{2}}(\hat{kr})^2/[1 + \frac{1}{4}\pi(kr)^2 \operatorname{Im} \Sigma_+],$$

$$(46) \quad (ka|\delta_{\min}|)^{\frac{1}{2}} = \frac{\pi}{\sqrt{2}}(kr)^2 \frac{(1 - \cos \theta_r)(1 - \cos \alpha)}{1 - 2 \cos \theta_r \cos \alpha},$$

which are in essential agreement with Twersky (1956, equation (117)).

The variation in intensity near an anomaly of the spectra associated with this particular grating should be qualitatively similar to Figs. 1 and 2, although the position of the minima will differ from those shown in the first figure. The maxima and minima are apparently much more sharp, and the ratio of intensities at maximum and minimum much greater than observed by Wood (1935) (for example, Wood mentions an intensity ratio of 15 to 20). This may be due to the fact that the present theory implies broad grooves and relatively narrow space between, while the gratings employed by Wood presumably had narrow grooves, with relatively wide spacing. Figures 1 and 2 do appear to indicate a tendency for the maxima and minima to broaden, and the intensity ratio to decrease, with increase in  $kr$ .

### 3. THE FINITE GRATING

Although of more practical interest than the preceding results, the solutions pertaining to a finite grating are severely complicated by end effects introduced through the non-periodic nature of the structure. It has been mentioned in I that the rigorous solution to the linear algebraic equations appears to be impracticable for large  $N$ , and at best an approximate solution may be expected.

It is unfortunate that the obtaining of even an approximation to the solution of the equations is made more difficult by the very property of the Hankel functions (discussed in the paragraph preceding equations (I,5)) which necessitated the present approach to the problem. Nevertheless, some progress can be made in the analysis, and the two boundary value problems, formulated previously, will be discussed more fully, attention being confined to small scatterers of elliptic section.

### 3.1. $u = 0$ on the Cylinders

The equations to consider are the integral equations (I,11) (which determine the distribution on the cylinders through equations (I,12)), and the linear algebraic equations (I,15). The former will be considered first; the index  $\nu$  runs from  $-N$  to  $N$  (rather than from 0 to  $N-1$ ), where  $N \gg 1$ . Thus the grating consists of  $2N+1$  elements.

It will be assumed that  $\xi_\lambda(\alpha)$  and  $G(\nu, \Lambda)$  possess expansions of the form

$$(47) \quad \begin{cases} b\xi_\lambda(\alpha)(\cosh^2 \mu - \cos^2 \eta_\lambda)^{\frac{1}{2}} = \sum_{n=0}^{\infty} \xi_\lambda^{(n)}(\alpha) \epsilon^n, \\ G(\nu, \Lambda) = \begin{cases} (2i/\pi)[\psi_0(\lambda, \Lambda) + \psi_1(\lambda, \Lambda) \epsilon^2 + \dots], & \nu = \lambda, \\ \sum_{n=0}^{\infty} G_n(\nu, \Lambda) \epsilon^n, & \nu \neq \lambda. \end{cases} \end{cases}$$

Here  $(\mu_\lambda, \eta_\lambda)$  represent elliptic co-ordinates with origin at the center of  $C_\lambda$  and  $\mu_\lambda = \mu$  (a constant) on each cylinder.  $\psi_0(\lambda, \Lambda)$  may be found by replacing  $\eta, \eta_0$  by  $\eta_\lambda, \eta_\Lambda$  respectively in the first of equations (I,52), while

$$(48) \quad G_0(\nu, \Lambda) = H(|\nu - \lambda|ka) - c(\nu - \lambda).$$

If the incident field also is expanded in powers of  $\epsilon$  it is found that the  $\xi_\lambda^{(0)}(\alpha)$  are the solution of the linear algebraic equations

$$(49) \quad \frac{2i}{\pi} (p + \mu) \xi_\lambda^{(0)} + \sum_{\nu \neq \lambda} [H(|\nu - \lambda|ka) - c(\nu - \lambda)] \xi_\nu^{(0)} = \frac{1}{2\pi} e^{i\lambda ka \cos \alpha},$$

$$\lambda = -N, -N+1, \dots, N-1, N.$$

Equations (49) are in form similar to equations (21†). In the latter, however, the coefficient of the unknown in the summation was merely  $H(|\nu - \lambda|ka)$ , of order  $|\nu - \lambda|^{-\frac{1}{2}}$ , while in the above equations, the coefficient is of order  $|\nu - \lambda|^{-3/2}$ . It is this characteristic which makes feasible the solution of equations (49) by iteration (successive substitutions) with  $\xi_\lambda^{(0)}(\alpha) = -\frac{1}{4}i(p + \mu)^{-1}e^{i\lambda ka \cos \alpha}$  as a first approximation. The series solution so obtained may be truncated after its first term provided  $ka$  is sufficiently large and, provided the factor  $H(|\nu - \lambda|ka) - c(\nu - \lambda)$  may be replaced by the first term in its asymptotic expansion, the process is convergent (uniformly in  $\alpha$ ) if

$$(50) \quad (ka)^{3/2} |p + \mu| \gtrsim \frac{1}{2} (\frac{1}{2}\pi)^{\frac{1}{2}} (1 - N^{-\frac{1}{2}}).$$

For instance, if  $\epsilon e^\mu = 0.01$ , this condition implies that  $ka \gtrsim 1.5$ , so the process is certainly convergent if  $ka$  is sufficiently large to justify use of the asymptotic expansion of  $H(|\nu - \lambda|ka)$ .

For the remainder of this discussion it will be assumed that  $\epsilon$  and  $ka$  are such that

$$(51) \quad \begin{cases} b\xi_\lambda(\alpha)(\cosh^2 \mu - \cos^2 \eta_\lambda)^{\frac{1}{2}} \doteq \xi_\lambda^{(0)}(\alpha), \\ \xi_\lambda^{(0)}(\alpha) \doteq -\frac{1}{4}i(p + \mu)^{-1}e^{i\lambda ka \cos \alpha}, \end{cases}$$

so to this order of approximation,  $-4i\xi_\lambda(\alpha)$  is equal to the normal derivative of the total wave function on a small, isolated cylinder  $C_\lambda$ .

To the same degree of approximation,

$$(52) \quad p_\lambda(\alpha, \theta) = 2\pi\xi_\lambda^{(0)}(\alpha),$$

and  $\sigma_\lambda(\alpha, \theta) \equiv \sigma_\lambda(\alpha)$ , which is independent of  $\theta$ , is determined by

$$(53) \quad \sigma_\lambda(\alpha) - \frac{1}{2}\pi i(p+\mu)^{-1} \sum_{\nu \neq \lambda} c(\nu-\lambda)\sigma_\nu(\alpha) = -8\pi i\xi_\lambda^{(0)}(\alpha),$$

$$\lambda = -N, -N+1, \dots, N-1, N.$$

Because of the slow decay of the coefficients  $c(\nu-\lambda)$ , a solution of equations (53) by the former iteration procedure appears to be convergent in  $N$ ,  $ka$ , and  $\alpha$ , only if interaction is unimportant, in which case the sum is small, and  $\sigma_\lambda = -8\pi i\xi_\lambda^{(0)}$  is a good first approximation. In order to obtain a solution by iteration valid when interaction is not negligible, it would appear essential to transform the equations to a system in which the off-diagonal terms of the coefficient matrix are more rapidly diminishing.

A more direct way to obtain an approximate solution is to separate the equations into blocks ( $M$  in number, say), in each of which the  $\sigma_\nu$  are assumed to differ by only the phase factor appropriate to the infinite grating, and to solve for the central unknown in each block. As  $M$  increases from unity, the approximation steadily improves, and yields the rigorous solution when  $M = 2N+1$ .

Twersky (1952) found it necessary to neglect end effects in order to proceed with his analysis, and essentially the same approximation ( $M = 1$ ) will be employed here. Then

$$(54) \quad \sigma_\nu(\alpha) \doteq \sigma_0(\alpha) e^{i\nu ka \cos \alpha},$$

where  $\sigma_0(\alpha)$  may be found from equations (53) on setting  $\lambda = 0$ . It is then apparent that  $\sigma_0(\alpha)$  is determined by the right-hand side of equation (I,56) (which pertains to the infinite grating), with  $\xi_0(\alpha)$  and  $\xi_0$  replaced by  $\xi_0^{(0)}(\alpha)$ , while the summation index in  $\sum_{\pm}$  runs from 1 to  $N$ , rather than from 1 to  $\infty$ . The approximation (54) should be best for small values of  $\nu$ , but undoubtedly the neglect of end effects will lead to some error in the field on the cylinders, and in the far zone to which region attention will henceforth be confined.

The field at a point  $P(R, \theta)$  (see Fig. 1 of I) is determined by equations (I,3) and (I,12). If the point of observation is sufficiently distant from the grating so that  $R \gg Na$ ,  $kR \gg 1$ , then  $H(kr_\nu)$  may be replaced by the first term in its asymptotic expansion, and

$$(55) \quad u(P) \sim u^1(P) - \frac{e^{i\pi\ell} e^{ikR}}{2\sqrt{(2\pi kR)}} \sum_{\nu=-N}^N e^{-i\nu ka \cos \theta} \sigma_\nu(\alpha, \theta),$$

where  $\sigma_\nu$  is defined by equations (I,10) or (I,13). To the order of approximation here considered,  $\sigma_\nu(\alpha, \theta)$  is independent of  $\theta$ , and equation (54) implies that the intensity of the scattered far field is equal to

$$(56) \quad |u^s(P)|^2 = \frac{|\sigma_0(\alpha)|^2}{8\pi kR} \left\{ \frac{\sin[(N+\frac{1}{2})ka(\cos\alpha - \cos\theta)]}{\sin[\frac{1}{2}ka(\cos\alpha - \cos\theta)]} \right\}^2.$$

It is evident that, to the present order of approximation, the variation in scattered intensity (other than due to the Fraunhofer factor in the curly brackets), will be much like that of the functions  $|a_\nu^{(0)}|^2$  which were discussed previously in connection with the infinite grating, while the intensities  $|a_\pi^{(0)}|^2$ ,  $|a_{-m}^{(0)}|^2$ , which refer to the grazing plane waves, correspond to  $\theta = 0$  and  $\pi$  respectively. For each of these values of  $\theta$ , the Fraunhofer factor attains its maximum of  $4(N+\frac{1}{2})^2$  at the corresponding Rayleigh wavelength.

Equation (56) is essentially equivalent to the result obtained in a different manner by Twersky (1952, equation (17)), whose analysis of the function  $\sigma_0(\alpha)$  predicts the existence and position of strong minima in the intensity of  $u^s$  (Wood anomalies). The reader is referred to this paper for a detailed discussion including several figures which illustrate the above phenomena. Minor differences between  $|a_\nu^{(0)}|^2$  and  $|\sigma_0(\alpha)|^2$  are apparent. The displacement of the position of the minimum from the Rayleigh wavelength is zero for the former case, and inversely proportional to  $N$  for the latter. Furthermore,  $|\sigma_0(\alpha)|^2$  (being of order  $N^{-1}$ ) does not drop to zero or possess a discontinuous slope at the minimum, while several subsidiary extrema, dependent on  $N$  and flanking the major anomaly become apparent.

### 3.2. $\partial u / \partial n = 0$ on the Cylinders

In this case, it is necessary to consider the integral equations (4) and the algebraic equations which may be derived from (6). If it is assumed that

$$(57) \quad \xi_\lambda(\alpha) = \sum_{n=0}^{\infty} \xi_\lambda^{(n)}(\alpha) \epsilon^n,$$

then, because on  $C$ , the normal derivative of  $G_0(\nu, \Lambda)$  vanishes if  $\nu \neq \lambda$  (see equations (47), (48)), equation (14) implies that  $\xi_\lambda^{(0)}(\alpha)$  is identical with the corresponding term for the infinite grating. The same is true for the function  $\xi_\lambda^{(1)}(\alpha)$ , so according to equations (15) and (17),

$$(58) \quad \begin{cases} \xi_\lambda^{(0)}(\alpha) = -\frac{1}{4}ie^{i\lambda ka \cos\alpha}, \\ \xi_\lambda^{(1)}(\alpha) = \frac{1}{4}e^\mu e^{i\lambda ka \cos\alpha} \cos(\eta_\lambda - \alpha + \beta). \end{cases}$$

Furthermore, the solution  $\xi_\lambda(\alpha)$  is, to order  $\epsilon$ , identical with the field determined in (†) (see equations (38†) for the scattered field). However, while in (†) it was assumed that interaction effects could be safely neglected, and equations (38†) gave the entire scattered field on the cylinders, here  $\xi_\lambda(\alpha)$  is but one part of the field on the scatterer, the total field being determined by equation (3).

The series expansion of  $q_\lambda(\alpha, \theta)$  (equation (7)) begins with the term  $i\epsilon^2 Q(\alpha, \theta) e^{i\lambda ka \cos\alpha}$ , where  $Q(\alpha, \theta)$  was introduced originally in equation (19) and is defined explicitly by equation (31). Thus, even to a first approximation,  $\tau_\lambda(\alpha, \theta)$  will be dependent on  $\theta$ , unlike the zero-order approximation to  $\sigma_\lambda(\alpha, \theta)$ .

If end effects are again neglected, and it is assumed that

$$(59) \quad \tau_\lambda(\alpha, \theta) = \tau_0(\alpha, \theta) e^{i\lambda k a \cos \alpha},$$

where  $\tau_0(\alpha, \theta)$  is determined from equation (6) with  $\lambda = 0$ , then several of the expressions pertaining to the infinite grating may, with minor modification, be employed here. For example,  $\tau_0(\alpha, 0)$  and  $\tau_0(\alpha, \pi)$  are determined by the right-hand sides of equations (18), if the subscript "0" is placed on the  $q$ -functions, and the summations involved in  $\sum_{\pm}$  run from 1 to  $N$ . Furthermore, if not both  $|\sum_{\pm}|$  are too large, and the functions  $q_k(\alpha, \theta)$  are replaced by the first term in their series expansions, then equation (25) gives the zero-order approximation  $[\tau_0^{(0)}(\alpha, \theta)]$  to  $\tau_0(\alpha, \theta)$ , in which end effects are neglected.

The far field of the grating is readily obtained from equation (1):

$$(60) \quad u(P) \sim u^1(P) + \frac{e^{i\pi t} e^{ikR}}{2\sqrt{(2\pi kR)}} \sum_{v=-N}^N e^{-ivka \cos \theta} \tau_v(\alpha, \theta).$$

To the present degree of approximation, the intensity of the scattered far field is equal to

$$(61) \quad |u^s(P)|^2 = \frac{|\tau_0^{(0)}(\alpha, \theta)|^2}{8\pi kR} \left\{ \frac{\sin[(N + \frac{1}{2})ka(\cos \alpha - \cos \theta)]}{\sin[\frac{1}{2}ka(\cos \alpha - \cos \theta)]} \right\}^2.$$

Thus, in this case also, the variation in scattered intensity (excluding the effects due to the Fraunhofer factor) will be similar to the behavior of the functions  $|b_{\nu}^{\pm(0)}|^2$ , while  $|b_n^{\pm(0)}|^2$  and  $|b_{-m}^{\pm(0)}|^2$  correspond to  $\theta = 0$  and  $\pi$ , respectively. Without doubt, the neglect of end effects, and the finite value of  $N$  will modify this behavior to some extent, but the presence of a maximum and a minimum of intensity on the long-wavelength side of an anomaly is to be expected (and was originally observed by Wood (1902) in the field scattered by a finite reflection grating).

#### ACKNOWLEDGMENTS

The author wishes to thank Mr. R. A. Hurd who read and commented upon the manuscript, and Mrs. M. Steen for assistance in the numerical calculations of Parts I and II.

#### REFERENCES

ARTMANN, K. 1942. Z. Physik. **119**, 529.  
 KARP, S. N. and RADLOW, J. 1956. IRE Trans. AP-4, 654.  
 MEECHAM, W. C. 1956. J. Appl. Phys. **27**, 361.  
 MILLAR, R. F. 1960. Can. J. Phys. **38**, 272.  
 PALMER, C. H. 1952. J. Opt. Soc. Am. **42**, 269.  
 TWERSKY, V. 1952. J. Appl. Phys. **23**, 1099.  
 ——— 1956. IRE Trans. AP-4, 330.  
 WOOD, R. W. 1902. Phil. Mag. Series 6, **4**, 396.  
 ——— 1912. Phil. Mag. Series 6, **23**, 310.  
 ——— 1935. Phys. Rev. **48**, 928.

# THE EFFECT OF QUENCHING AND NEUTRON IRRADIATION ON INTERNAL FRICTION OF ALUMINUM-5% MAGNESIUM ALLOY<sup>1</sup>

W. G. NILSON

## ABSTRACT

Low frequency internal friction data were obtained on the alloy aluminum-5% magnesium to determine the effects of quenching and neutron irradiation on solute movement. A damping peak near 150° C, attributed to stress-induced reorientation of solute atoms, was found to be shifted to lower temperatures by these treatments. This behavior corresponds to a reduction in relaxation time for the damping process, and is compatible with the idea that the solute atoms act as traps for vacancies.

## INTRODUCTION

It has been suggested that in substitutional solid solutions, solute atoms can act as traps for vacancies when their atomic size is appreciably different from the solvent (Seitz 1950). An excess concentration of these defects may enhance solute atom diffusion. Recent work used this argument to explain certain phenomena which occurred during recovery, precipitation, and strain ageing of binary aluminum-magnesium alloys (Perryman 1955, 1956; Westwood and Broom 1957).

To study the effect of an excess concentration of vacancies on diffusion it is necessary that the experiments be conducted at temperatures where rapid annealing out of the defects does not occur. Conventional diffusion experiments at elevated temperatures are therefore not suitable. Atomic mobility may be studied at lower temperatures by utilizing the anelastic properties of certain solid solutions (Nowick 1952a). This method is based upon the measurement of the relaxation time associated with stress-induced reorientation of solute atoms. An anelastic relaxation can be examined experimentally by observations on the damping capacity of a material. Although the exact nature of the local ordering is somewhat uncertain (Zener 1947; LeClaire and Lomer 1954) it seems necessary that the atomic diameters of solvent and solute be substantially different. In binary aluminum-magnesium alloys, an internal friction peak appears between 100° and 200° C at a vibration frequency of the order of 1 cycle per second (Berry 1956; Maringer *et al.* 1956), and the magnitude of the peak is dependent on solute concentration. For a fixed frequency, quenching displaces the peak maximum to lower temperature, thus indicating a reduction in the relaxation time. These effects are characteristic of a relaxation produced by stress-induced local ordering (Berry and Nowick 1958; Li and Nowick 1956; Nowick 1951; Nowick and Sladek 1953; Roswell and Nowick 1953). It appeared, therefore, that a binary

<sup>1</sup>Manuscript received April 5, 1960.

Contribution from the Research Laboratory, The International Nickel Company, Inc., Bayonne, N.J. The work described was done at the Chalk River laboratories of Atomic Energy of Canada Limited. This is a revised version of the original report AECL-701 published at Chalk River, September, 1958.

Issued as A.E.C.L. No. 701.

aluminum-magnesium alloy would be a suitable material on which to study the effect of excess vacancy concentration on atomic mobility. Such defects can be produced by irradiation with fast neutrons as well as by quenching. Inasmuch as the radioactivity of aluminum and magnesium resulting from irradiation with pile neutrons is low, the use of such an alloy was particularly attractive.

## EXPERIMENTAL

### *Torsion Pendulum*

Internal friction data were obtained as a function of temperature. The low frequency torsion pendulum used was similar to that described by Ke (1947). The wire specimens were 6 in. long and 0.025 in. diameter. The total tensile load on the specimens was 350 p.s.i. The vibration amplitude used was such that the maximum shear strain at the specimen surface was  $1.3 \times 10^{-5}$ . Measurements on a few specimens at room temperature using maximum shear strains somewhat above and below this value showed no appreciable amplitude dependence of the damping.

The temperature variation along the specimen length was maintained within approximately plus or minus 1°C at 200°C. The thermal inertia of the furnace allowed heating rates of 5°C per minute. The free cooling rate was exponential with a decay time of about 20 minutes.

### *Material Used*

High purity aluminum-5% magnesium alloy in the form of hot-rolled slab was drawn to wire with a final diameter of 0.025 in. Spectrographic analysis gave the approximate impurity content as 0.03% silicon, 0.004% copper, and less than 0.005% iron. After they were cut into lengths of about 6 in. all specimens were annealed at 500°C for 4 hours for straightening and also to produce a large grain size. A large grain size was desired to suppress a high background that might occur from grain boundary relaxation (Nowick 1952b). The resultant average grain diameter was of the same order of magnitude as the specimen diameter.

Specimens were usually heat-treated in argon. However, in preliminary tests using air heat treatment, the surface oxide so produced did not seem to have any obvious effect on the damping. At 500°C the temperature variation along the specimen length in the annealing furnace was about plus or minus 2°. Quenching was accomplished by blowing the specimen rapidly into a water bath at room temperature.

### *Neutron Irradiation*

The irradiations were carried out at The Atomic Energy of Canada, Ltd. (Chalk River, Ontario) in NRX reactor fuel rod positions using a transformer rod similar to that described by Cook and Cushing (1953). Air cooling maintained the temperature of the specimens at about 60°C, as measured with iron-constantan thermocouples. Calculations indicate that within the irradiation assembly the neutron flux above 500 ev is about  $3.1 \times 10^{13}$  n/cm<sup>2</sup> sec with the reactor operating at 40 megawatts.

Prior to irradiation the 6-in. long wire specimens were given various heat treatments, described later. After removal from the reactor and before making the damping tests the specimens were stored for several weeks at room temperature. The resultant activity of the wires was then sufficiently low to permit direct handling with rubber gloves when the torsion pendulum was being loaded.

## RESULTS

### Effect of Quenching

Internal friction data was taken at numerous temperatures on *heating* from room temperature to almost 200° C. Tests were carried out on specimens with thermal and irradiation treatments listed in Table I. The initial anneal of 500° C for 4 hours preceded all those shown.

TABLE I  
Specimen treatments

Specimen	400° C, 1/2 hr	200° C, 1/2 hr	60° C	Irradiation, n.v.t. (at 60° C)	Storage conditions	Notes
A	Slow FC <sup>a</sup>	—	—	—	—	1
B	WQ <sup>b</sup>	—	—	—	—	
C	WQ	—	—	—	—	2
D	WQ	—	—	—	—	2
F	WQ	—	15 days	—	—	
H	WQ	WQ	—	—	—	
I	WQ	Slow FC	—	—	—	1
J	WQ	Fast FC	60 days	—	-40° C, 28 days	3
K	WQ	—	—	$4.4 \times 10^{19}$	—	
M	WQ	—	—	$1.1 \times 10^{20}$	—	4
N	WQ	Fast FC	—	$1.1 \times 10^{20}$	-40° C, 29 days	3
O	WQ	Fast FC	—	$1.1 \times 10^{20}$	-40° C, 21 days	3, 5
P	WQ	Fast FC	—	$1.1 \times 10^{20}$	-40° C, 48 days	2, 3

NOTES: <sup>a</sup>FC, furnace cooling.

<sup>b</sup>WQ, water quenching.

1. Slow FC is exponential to room temperature with 2 hours decay time.

2. Used for isothermal annealing tests near 100° C.

3. Fast FC is exponential to room temperature with 30-minute decay time.

4. Chemical pickling for 3 minutes at 80° C occurred prior to irradiation.

5. Data confirms results with specimen N.

The internal friction curve for a specimen which had been slowly cooled from 400° C is shown in Fig. 1. On first heating to 200° C the damping below 90° C was relatively high. This behavior was not seen on repeating the run. In both cases, however, a peak maximum was seen at about 150° C. With a fast water quench from 400° C, a broad maximum was seen at about 100° C, plus a smaller one at 150° C, Fig. 2. On retesting a second and third time, a stable maximum at 150° C was observed. Quenching from 400° C followed by ageing at 60° C for 15 days resulted in a broad maximum at about 125° C, plus a lower one at 150° C, Fig. 3. A second run with such a specimen showed a stable maximum at 150° C similar to that shown in Fig. 2. It appeared that quenching had the effect of either shifting the stable peak or introducing a peak at lower temperatures. Ageing at 60° C resulted in displacement of the low temperature peak towards the stable temperature (150° C), while a few

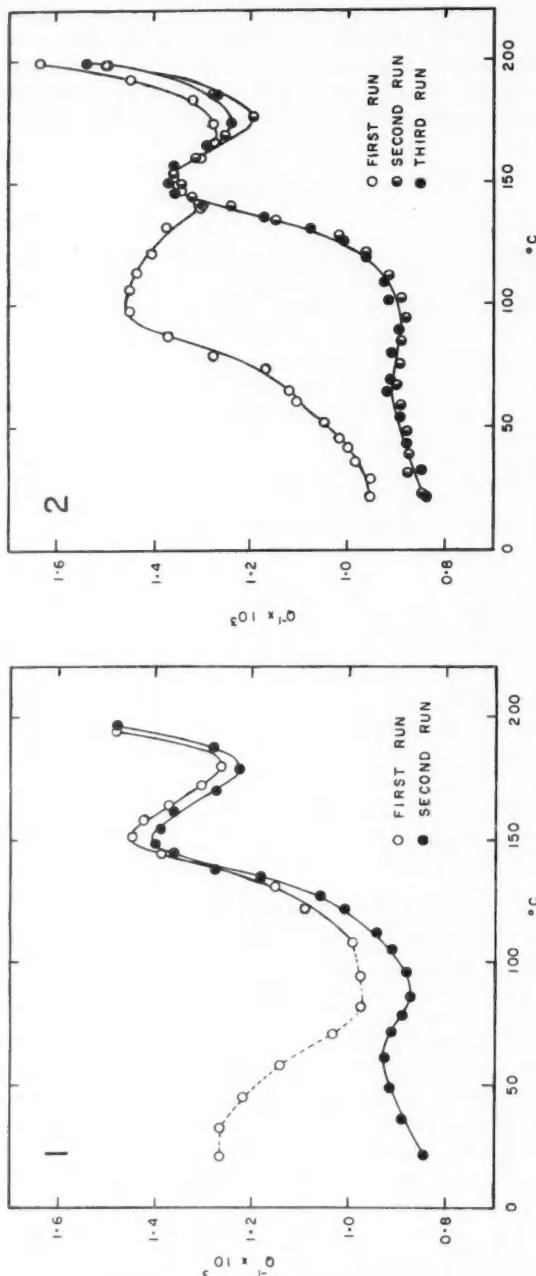


FIG. 1. Internal friction vs. temperature. Specimen A, 400° C, 1 hr, slow FC.  
 FIG. 2. Internal friction vs. temperature. Specimen B, 400° C,  $\frac{1}{2}$  hr, WQ.

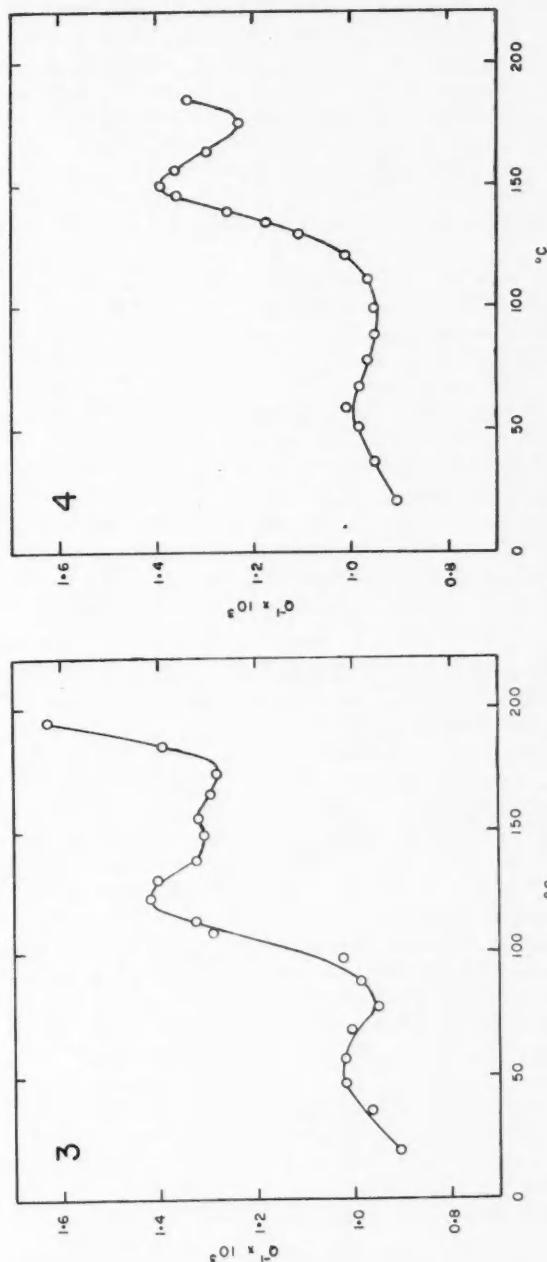


FIG. 3. Internal friction vs. temperature. Specimen F,  $400^{\circ}\text{C}$ ,  $\frac{1}{2}$  hr; WQ +  $60^{\circ}\text{C}$ , 15 days.  
 FIG. 4. Internal friction vs. temperature. Specimen H,  $400^{\circ}\text{C}$ ,  $\frac{1}{2}$  hr; WQ +  $200^{\circ}\text{C}$ ,  $\frac{1}{2}$  hr, WQ.

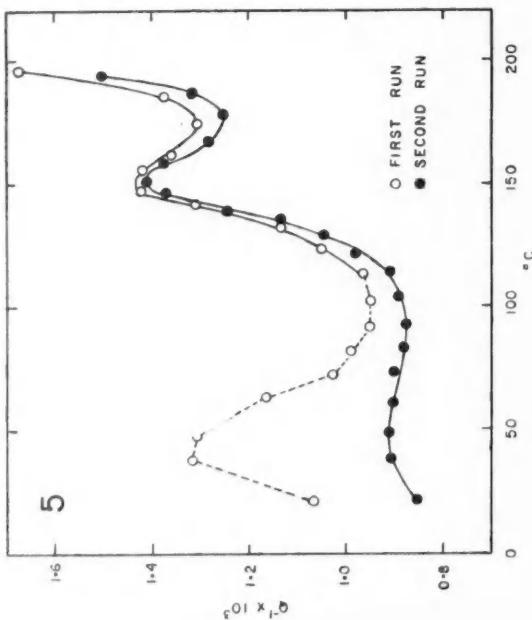
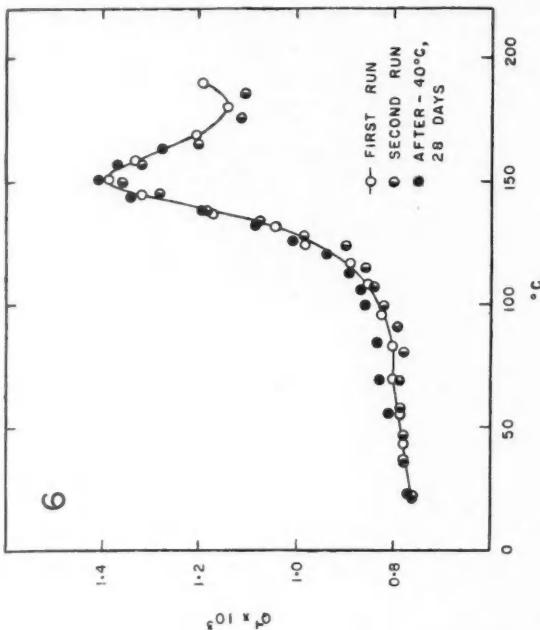


FIG. 5. Internal friction vs. temperature. Specimen I,  $400^\circ\text{C}$ ,  $\frac{1}{2}$  hr;  $WQ+200^\circ\text{C}$ ,  $\frac{1}{2}$  hr, slow FC.  
 FIG. 6. Internal friction vs. temperature. Specimen J,  $400^\circ\text{C}$ ,  $\frac{1}{2}$  hr;  $WQ+200^\circ\text{C}$ ,  $\frac{1}{2}$  hr, fast FC+ $60^\circ\text{C}$ , 60 days; remounted after  $-40^\circ\text{C}$ , 28 days.

minutes near 200° produced complete stability. This suggested that only a single peak actually occurred; quenching resulted in its shift to lower temperatures, and annealing after quenching caused its migration towards a stable temperature. An explanation of the appearance of two peaks in a single experimental run, as in the first heating data of Figs. 2 and 3 would be that annealing took place during the test. The time taken in going from 100° to 150° C was about 1 hour. The difference between peak heights at lower and higher temperatures could be mostly accounted for in differences of background damping. Isothermal annealing experiments are described later which show this peak migration.

It was anticipated, therefore, that quenching from 400° C followed by annealing at 200° C and subsequent quenching would give only a stable damping peak, and this is shown in Fig. 4. However, when a sample was slowly cooled from 200° C, then in addition to the stable peak a relatively high damping appeared below 90° C, Fig. 5, similar to that of Fig. 1 for furnace cooling from 400° C. It seemed that this high damping just above room temperature resulted only with very slow cooling below 200° C. It was not seen after a specimen was quenched or even allowed to cool in the torsion pendulum from 200° C. A fast furnace cool from 200° C, followed by ageing at 60° C for 60 days produced the stable peak, Fig. 6, and a very small peak barely resolvable in the vicinity of 50°–70° C. In all tests this small peak seemed quite broad; the magnitude varied somewhat among the tests, but not in a manner that was readily related to prior treatment.

#### *Effect of Neutron Irradiation*

A specimen quenched from 400° C and exposed to an irradiation of  $4.4 \times 10^{19}$  n.v.t. showed on first heating a peak maximum at about 135° C. On reheating, it appeared in the usual stable position at 150° C. The position of the maximum on the first heating of this quenched and irradiated specimen was at a higher temperature (135° C) than for quenching alone (100° C). However, since the irradiation temperature was in the vicinity of 60° C, the effect of irradiation without the concurrent thermal ageing was uncertain. Ageing a quenched and unirradiated specimen at 60° C for about the same time also resulted in the peak being closer to the stable position, Fig. 3.

A quenched specimen irradiated for  $1.1 \times 10^{20}$  n.v.t. showed a peak on first reheating at about 130° C. The increase in irradiation exposure would have been expected to move the peak maximum towards its stable position at 150° C due to thermal ageing alone. This was not the case, however, for the effect, if any, of the longer irradiation was to move the peak to a slightly lower temperature. This suggested the possibility that defects were being introduced by irradiation and retained at 60° C.

In an attempt to observe the effects of irradiation alone, specimens were exposed after quenching from 400° C plus fast furnace cooling from 200° C. Such a specimen if unirradiated and aged at 60° C produced only a stable peak, Fig. 6. However, with an irradiation of  $1.1 \times 10^{20}$  n.v.t. the results from two specimens showed a peak maximum at about 135° C rather than 150° C on first reheating, Fig. 7.

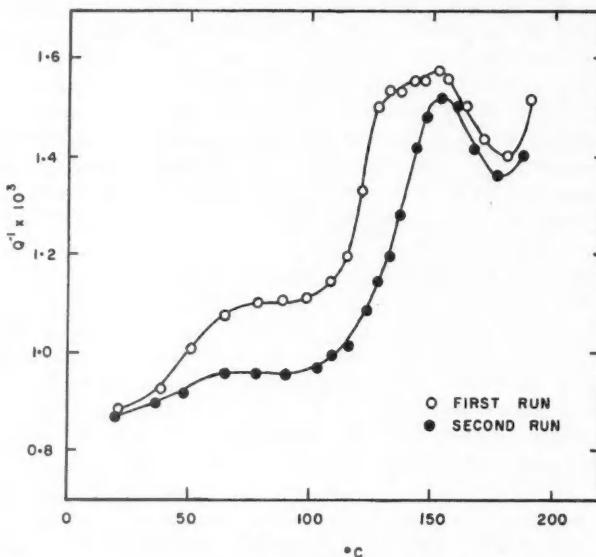


FIG. 7. Internal friction vs. temperature. Specimen N, 400°C,  $\frac{1}{2}$  hr; WQ+200°C,  $\frac{1}{2}$  hr; fast FC+neutron irradiation  $1.1 \times 10^{20}$  n.v.t.

An explanation for the marked lack of symmetry of the peak of the first run in Fig. 7 would be that annealing during the test has caused the peak to shift. About  $\frac{1}{2}$  hour elapsed while data between 130° and 150° C were being taken.

It appears, therefore, that both quenching and irradiation displace the internal friction peak from its stable position to lower temperatures. Such a shift implies that these treatments decrease the relaxation time characteristic of the process responsible for the peak.

The height of the damping peak in any individual specimen (after subtracting the room temperature background value) did not seem to be influenced by ageing during a test. Its magnitude among all specimens was usually in the range  $0.5\text{--}0.7 \times 10^{-3}$  in  $Q^{-1}$  units. Variations within this range did not appear to be related to the various ageing and irradiation treatments.

#### Activation Energy for Stable Peak

Two methods were used for determining the activation energy associated with the stable internal friction peak.

The first method was based on the observed temperature change for the peak maximum when the frequency was changed. Increasing the vibration frequency by a factor of about  $2\frac{1}{3}$  resulted in an increase of the stable peak maximum temperature from 150° C to about 160° C. It was then assumed that the relaxation time follows an Arrhenius type of relationship,

$$(1) \quad \tau = \tau_0 \exp(H/RT)$$

where  $\tau$  is the mean relaxation time,

$H$  is the activation energy.

From the temperature shift of the peak when the frequency is changed the activation energy was calculated. Since there seemed to be no distinct separation between values for unirradiated and irradiated specimens, or even any marked effect of the prior treatments used, an average of all the results gave 27.7 kcal/mole with a mean deviation of 1.3 kcal/mole. This is in agreement with the value of 27.8 kcal/mole previously reported (Maringer *et al.* 1956) for this peak in an alloy of very similar composition.

The second method used for determining the activation energy from data on the stable peak was to analyze the shape of the peak. For low damping the internal friction is given by (Zener 1948; Nowick 1953)

$$(2) \quad (Q^{-1} \text{ observed}) - (Q^{-1} \text{ background}) = \tan \phi = \Delta \{\omega \tau / [1 + (\omega \tau)^2]\}$$

where  $\phi$  is the phase difference between stress and strain,

$\omega$  is the angular speed of vibration,

$\Delta$  is the relaxation strength; in this case it is twice the peak maximum  $\tan \phi$ .

Solving equation (2) for  $\tau$  gives

$$(3) \quad \tau = (1/\omega) \cdot [Z + \sqrt{(Z^2 - 1)}]$$

where

$$Z = \Delta / (2 \cdot \tan \phi)$$

Equation (3) is valid only when the process has a unique relaxation time (at any particular temperature). Making this assumption the equivalent  $\tau$  was calculated for measured values of the internal friction from which were subtracted the estimated background damping. The peak maximum corresponds to  $\omega \tau = 1$ . Damping values at higher temperatures correspond to  $\omega \tau < 1$ , while at lower temperatures  $\omega \tau > 1$ . To reduce the error introduced by the rising background at high temperatures only the low temperature side of the stable peak was analyzed. Values of  $\log \tau$  so determined were plotted against the reciprocal of the corresponding absolute temperatures. The activation energies were then calculated from the almost linear slopes of these lines. Again there seemed to be no marked differences between irradiated, unirradiated, and the variously heat-treated specimens. The results gave an average of 24.0 kcal/mole with a mean deviation of 1.8 kcal/mole.

The activation energy from the peak shape analysis is lower than that determined by the frequency shift method. This indicates that the peak is broader than it should be for unique values of  $\tau$ . This is a characteristic of this type of relaxation and can be ascribed to a distribution of relaxation times (Nowick 1953; Berry and Nowick 1958).

The difference between activation energies determined by peak shape analysis and by the frequency shift method signify that equation (2) is, therefore, not valid for the shape of the stable peak.

### Annealing of the Peak Instability

Several attempts were made with quenched and with irradiated specimens to follow the isothermal change in damping as the peak moved from the unstable position to its stable position at about 150° C. With a quenched specimen the unstable peak maximum was near 100° C, Fig. 2. Isothermal data were taken on such specimens by heating them quickly to temperatures between 100° and 150° C and noting the subsequent change of damping with time. If the 100° C peak was a second one in addition to the stable maximum at 150° C, then the isothermal damping at intermediate temperature should initially decrease. Such a decrease would indicate that the low temperature peak maximum was diminishing without necessarily changing temperature. However, data taken with quenched specimens at 111° C and 132° C show an initial increase to a maximum, followed by a decrease, Fig. 8. This indicates that the 100° C peak is moving upwards in temperature during the anneal.

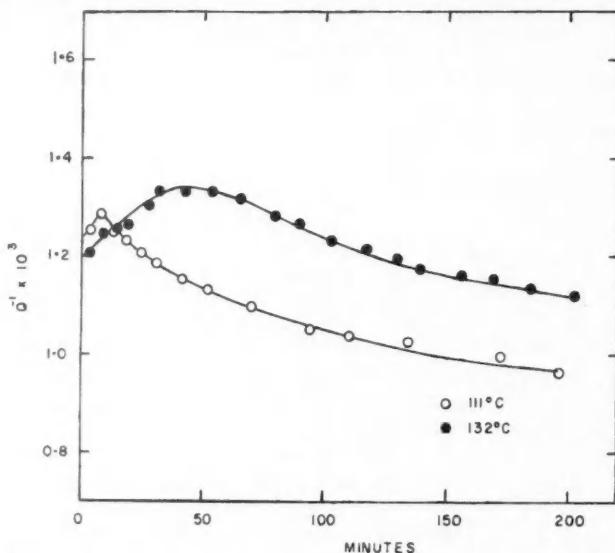


FIG. 8. Internal friction vs. annealing time. Specimen C (111° C data), 400° C,  $\frac{1}{2}$  hr; WQ. Specimen D (132° C data), 400° C,  $\frac{1}{2}$  hr, WQ.

The initial damping values of Fig. 8 are on the high temperature side of the original peak, corresponding to  $\omega\tau < 1$ . The maxima correspond to  $\omega\tau = 1$ . Annealing thus resulted in a gradual increase in the relaxation time of the quenched material. The explanation for the appearance of two peaks in the first run of Figs. 2 and 3 is, therefore, that such annealing took place during the experiments.

Data were also taken with an irradiated specimen. Irradiation had shifted the peak to lower temperatures, as in Fig. 7. Subsequent isothermal holding at 132° C produced the peak shift indicated by Fig. 9.

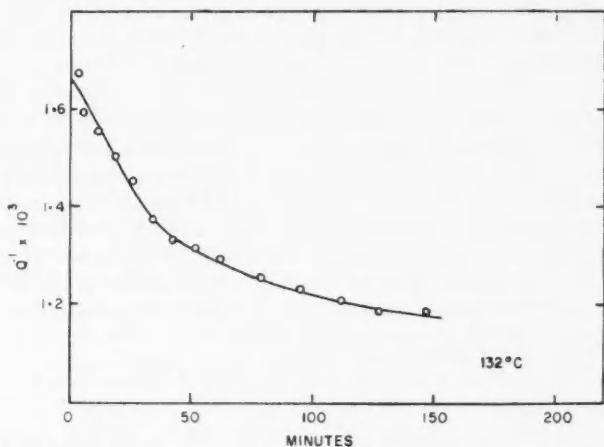


FIG. 9. Internal friction vs. annealing time. Specimen P, 400° C,  $\frac{1}{2}$  hr; WQ + 200° C,  $\frac{1}{2}$  hr; fast FC + neutron irradiation  $1.1 \times 10^{20}$  n.v.t.

## DISCUSSION

### *Effect of Quenching*

The effect of quenched-in vacancies on reducing the relaxation time for the stress-induced reorientation of solute atoms has been observed by others as mentioned. As a result of this the internal friction peak is shifted to lower temperatures providing the defects do not anneal out rapidly at room temperature. This has been demonstrated in the present case for aluminum-magnesium alloy where annealing was required to remove the quenching effect.

Work by Panseri and co-workers (1958) on the effects of quenching aluminum-magnesium alloys with up to 1.7% magnesium shows that the presence of the solute atoms inhibits the recovery of excess electrical resistivity at room temperature. A recovery stage occurred at 100°-150° C. From their results, as well as those of Westwood and Broom (1957) on strain aging, they suggest that at room temperature solute atoms migrate to dislocations, and that quenching promotes this movement through the introduction of vacancies. The recovery of resistivity in the vicinity of 100° C was then ascribed to a release of the entrapped excess vacancies from solute atoms.

The present internal friction results are compatible with this view. The reduction in relaxation time for stress-induced solute movement by quenching thus appears to be due to increased solute mobility when associated with a higher than equilibrium concentration of vacancies. Annealing at 100°-150° C to remove the excess vacancies resulted in a decrease in solute mobility and consequent increase in the relaxation time. This was indicated by the damping peak shift of quenched material from lower to higher temperatures during such annealing.

Maringer and co-workers (1956) have given a different interpretation to the shift in damping peak by quenching. They observed that quenching an aluminum-5½% magnesium alloy specimen from 500° C produced, on heating,

a low frequency internal friction peak with maximum at about 100° C; after ageing at 250° C for 18 hours it occurred at 160° C. These authors suggested that the temperature change was due to magnesium atoms migrating to dislocation sites where stress-induced movement required a greater relaxation time. However, Panseri and co-workers (1958) suggested that in quenched material the movement of solute atoms to dislocations occurred rapidly at room temperature. In the present internal friction experiments the specimens had been aged for comparatively long times at room temperature prior to measurement and, therefore, the magnesium atoms were presumably already at dislocation sites. It is likely, therefore, that the shift observed by Maringer *et al.* (1956) may be the same as observed in the present case where it is attributed to the effect of excess vacancies on increasing solute mobility.

#### *Effect of Neutron Irradiation*

Inasmuch as neutron irradiation can also induce lattice defects, the peak shift to lower temperatures after irradiation was not entirely unexpected. It is noteworthy, however, that the effect did not anneal out during irradiation or long storage at room temperature. No effect of irradiation was found by Li and Nowick (1956) on solute mobility in an alloy of 17 at.-% aluminum in copper. In that material the relaxation strength was almost an order of magnitude greater than for the aluminum-5% magnesium alloy. Those authors were thus able to use elastic after-effect measurements for direct determinations of quenching and irradiation effects on the relaxation time. In the present experiments with aluminum-magnesium the after-effect method would have been insensitive to the relaxation. Li and Nowick (1956) found that quenching markedly increased solute mobility. However, no effect was found after neutron irradiation to an integrated neutron flux of about  $10^{19}$  n.v.t. No details of the fast flux relative to the thermal flux were given. Specimens were irradiated at 50° C, 35° C, and at liquid nitrogen temperatures, but were tested at and above room temperature. These authors attributed the negative results to the annihilation of vacancies by the more mobile interstitials also produced by the irradiation.

In the present case the effect is still very much less than that of quenching; compare Fig. 7 with Fig. 2. This is possibly a result of thermal ageing at the temperature of irradiation, 60° C, as well as defect annihilation by interstitial-vacancy combination.

#### *Damping Peak near Room Temperature*

In two instances a relatively large damping was found near room temperature (Figs. 1 and 5). In both cases the prior heat treatment was one of slow furnace cooling from at least 200° C to room temperature (20°-25° C). For the specimens cooled from 200° C three measurements were initially taken at room temperature. The internal friction decreased with each succeeding measurement. Unfortunately, this was not done on the specimen cooled from 400° C to show whether the same trend would occur. Upon completing a test with each specimen, this high damping disappeared. Only a small peak remained, as was seen in other tests. The nature of this damping was not

investigated. It is conceivable, however, that it may have originated from some type of stress sensitive solute segregation. Damask and Nowick (1955) have suggested that applied stress may interact with stresses that exist in regions of coherency between matrix and precipitate.

Inasmuch as the magnesium concentration was above the solid solubility limit, some segregation may have occurred at any early stage of precipitation. Fast cooling suppressed this while slow cooling promoted it. Its disappearance after a single heating to 200° C followed by the fast cooling in the torsion pendulum might have been due to a re-resolution of whatever preferential arrangement had occurred. Geisler *et al.* (1943), LaCombe (1944), and Perryman (1951) have indicated that at such lower temperatures the formation of  $\beta$  phase ( $\text{Al}_3\text{Mg}_2$ ) is preceded by an unstable transition structure. The observed high damping might have been a result of a relaxation originating in regions between matrix and segregates in an early stage of precipitation.

#### SUMMARY

1. The experimental information for the low frequency internal friction peak near 150° C in an aluminum-5% magnesium alloy indicates that it is likely due to stress-induced local ordering of the solute atoms. Its observed movement to lower temperatures after either quenching or neutron irradiation, corresponding to a reduction in relaxation time for the damping process, is compatible with the idea that the magnesium atoms act as traps for the vacancies produced by these treatments.
2. The annealing out of the damping effects produced by quenching and irradiation occurred around 100° C. This corresponds to the recovery of quenched-in excess electrical resistivity attributed to the release of the defects from solute atoms (Panseri *et al.* 1958).
3. The activation energy of 27.7 kcal/mole determined for the stable peak was in agreement with internal friction results of others (Maringer *et al.* 1956).
4. A high damping observed near room temperature in slowly cooled material may be due to a stress sensitive solute segregation, for example, some change in structure in the very early stages of precipitation.

#### ACKNOWLEDGMENTS

The author wishes to acknowledge the use of the Research Metallurgy facilities at The Atomic Energy of Canada, Ltd. (Chalk River, Ontario) where this work was done.

Particular thanks are made to Drs. B. S. Berry and A. S. Nowick in discussing some of their preliminary internal friction work with aluminum-magnesium alloys.

#### REFERENCES

BERRY, B. S. 1956. Private communication (Yale Univ.).  
BERRY, B. S. and NOWICK, A. S. 1958. NACA-TN-4225.  
COOK, L. G. and CUSHING, R. L. 1953. *Acta Met.* **1**, 539.  
DAMASK, A. C. and NOWICK, A. S. 1955. *J. Appl. Phys.* **26**, 1165.

GEISLER, A. H., BARRETT, C. S., and MEHL, R. F. 1943. *Trans. AIME*, **152**, 201.  
KE, T. S. 1947. *Phys. Rev.* **71**, 533.  
LA COMBE, P. 1944. *Rev. Met.* **41**, 180, 217, 259.  
LECLAIRE, A. D. and LOMER, W. M. 1954. *Acta Met.* **2**, 731.  
LI, C. Y. and NOWICK, A. S. 1956. *Phys. Rev.* **103**, 294.  
MARINGER, R. E., MARSH, L. L., and MANNING, G. K. 1956. NACA-TN-3681.  
NOWICK, A. S. 1951. *Phys. Rev.* **82**, 551.  
——— 1952a. *Phys. Rev.* **88**, 925.  
——— 1952b. *Metal interfaces* (American Society for Metals), 248.  
——— 1953. *Progr. in Metal Phys.* **4**, 1.  
NOWICK, A. S. and SLADEK, R. J. 1953. *Acta Met.* **1**, 131.  
PANSERI, C., GATTO, F., and FEDERIGHI, T. 1958. *Acta Met.* **6**, 198.  
PERRYMAN, E. C. W. 1955. *Acta Met.* **3**, 412.  
——— 1956. *J. Metals*, **8**, 1247.  
PERRYMAN, E. C. W. and BROOK, G. B. 1951. *J. Inst. Metals*, **79**, 19.  
ROSWELL, A. E. and NOWICK, A. S. 1953. *J. Metals*, **5**, 1259.  
SEITZ, F. 1950. *Acta Cryst.* **3**, 355.  
WESTWOOD, A. R. C. and BROOM, T. 1957. *Acta Met.* **5**, 249.  
ZENER, C. 1947. *Phys. Rev.* **71**, 34.  
——— 1948. *Elasticity and anelasticity of metals* (Chicago).

# TURBULENT DIFFUSION IN TWO AND THREE DIMENSIONS BY THE RANDOM-WALK MODEL WITH MEMORY<sup>1</sup>

R. C. BOURRET

## ABSTRACT

A lattice model used for the derivation of the telegraph equation for diffusion is extended to two and three dimensions. Appropriate generalizations of the telegraph equation are obtained. These equations give a fine-grained chronological description of diffusion. From these equations, the velocity autocorrelation functions of the diffusing particles are obtained.

## I. INTRODUCTION

This paper is one of a series dealing with the problem of establishing the equations of turbulent diffusion. By turbulent diffusion is meant diffusion in which the movement of the diffusing particles is not completely incoherent. This implies that the velocity autocorrelation functions of the particles, or tagged fluid elements, have finite values for finite correlation times. It is possible to assume the form of these velocity autocorrelation functions and then proceed to seek appropriate diffusion equations embodying them as characterizations of the diffusive process. This procedure is followed in other papers of this series (see, for example, Bourret 1960).

In this article, we wish to apply the model of the random walk in a lattice in which the direction of the random step is a stochastic function of the direction of the previous step taken. This means that although the actual direction chosen is a statistical matter, the relative probabilities of the various directions are determined by the previous direction. In this sense memory is operative. This model was first employed by Taylor (1921) and more fully developed by Goldstein (1951) for the case of one dimension. It has been further investigated by Gupta (1959), Michelson (1954), and Davies *et al.* (1954). The last two authors mentioned state the two-dimensional problem but do not solve it. Here we shall derive the two- and three-dimensional diffusion equations implied by the use of the Taylor-Goldstein model.

The procedure used is to employ a two-, and later a three-, dimensional lattice model, or space, in which particles are imagined hopping from site to site according to some stochastic rule. A simple description of this process gives a set of finite difference equations. These are then reduced to partial differential equations by making the lattice infinitely fine-grained. A few auxiliary quantities are defined in making the transition to the continuous lattice. The lattice is the scaffold on which the quantitative description is erected, but the physical process imagined must be independent of the scaffold. The transition to a continuous lattice removes part of the scaffold's effect upon the resulting description, but not all. The "warp" and the "woof", i.e., the

<sup>1</sup>Manuscript received June 14, 1960.

Contribution from the Hughes Research Laboratories, Hughes Aircraft Company, Malibu, California.

directions of the lattice symmetry axes, remain. If the resulting diffusion equations are to be applied to isotropic diffusion, and not to diffusion in a crystalline or otherwise anisotropic medium, the favored axes must be suppressed. This is simply accomplished by averaging the anisotropic differential expressions over all directions. This procedure is equivalent to averaging over all scaffold or lattice orientations in the underlying description. In practice, as will be seen, this proves easy to accomplish.

## II. THE TWO-DIMENSIONAL CASE

Figure 1 shows the lattice envisioned. Diffusing particles may hop from site to site, one at a time, along the lattice axes. For convenience in notation, the directions will be designated by  $N$ (north),  $S$ (south),  $E$ (east), and  $W$ (west).

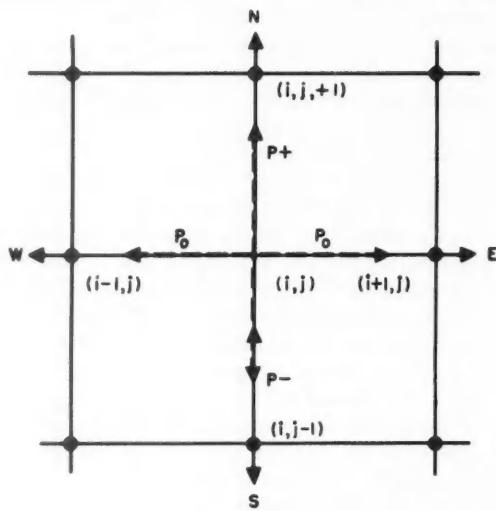


FIG. 1.

The diffusing particles are divided into four classes, according to the direction in which they made their last transition. Thus, at the site  $(i, j)$  there may be  $N_k^{i,j}$  particles arrived by a northward step,  $S_k^{i,j}$  by a southward step, and so on for  $E_k^{i,j}$  and  $W_k^{i,j}$ . The superscripts indicate location and the discrete time variable  $k$  is shown as a subscript. Time is measured in units  $\tau$  and distance in units  $\lambda$ . If a particle is at, say, site  $(i, j)$  it may make its next hop in any of four directions. The probabilities of these four alternatives are not generally equal, nor are they weighted in terms of absolute directions. They are weighted in a fixed way in the co-ordinates determined by the direction of the preceding hop. Thus, they may be designated by:  $p_+$ ,  $p_-$ , and  $p_0$ , representing the probability for a step in the same direction as in the previous step, a reversal of direction, or a turn to the right or the left. The left- and right-hand turn

probabilities may readily be chosen to be different, but we take them to be equal here, although if a description of diffusion in currents influenced by magnetic or Coriolis forces were contemplated, this refinement might be worth making.

Since the particles at any one site must have arrived from one of the four adjoining sites where they were located at the previous time, we may write the following equations from an examination of the figure:

$$(1) \quad \begin{aligned} N_k^{i,j} &= p_+ N_{k-1}^{i,j-1} + p_- S_{k-1}^{i,j-1} + p_0 E_{k-1}^{i,j-1} + p_0 W_{k-1}^{i,j-1}, \\ S_k^{i,j} &= p_- N_{k-1}^{i,j+1} + p_+ S_{k-1}^{i,j+1} + p_0 E_{k-1}^{i,j+1} + p_0 W_{k-1}^{i,j+1}, \\ E_k^{i,j} &= p_0 N_{k-1}^{i-1,j} + p_0 S_{k-1}^{i-1,j} + p_+ E_{k-1}^{i-1,j} + p_- W_{k-1}^{i-1,j}, \\ W_k^{i,j} &= p_0 N_{k-1}^{i+1,j} + p_0 S_{k-1}^{i+1,j} + p_- E_{k-1}^{i+1,j} + p_+ W_{k-1}^{i+1,j}. \end{aligned}$$

These equations, together with a specified set of probabilities  $p_+$ ,  $p_-$ , and  $p_0$ , completely determine, along with boundary conditions, the nature of the diffusion in the lattice. Since the site-to-site transition probabilities are determined by the direction of the immediately preceding transition, it is seen that the process is Markovian. The velocity autocorrelations, however, prove to be sums of simple exponentials rather than single exponential functions. The more usual result that the correlation functions are simple exponentials is true only in one dimension.

The next step is the reduction of equations (1) to differential equations by explicitly introducing the distance and time intervals  $\lambda$  and  $\tau$ , expanding to first order in the space and time intervals, and dividing by the time interval. This process yields the following set of differential equations in the limit  $\tau \rightarrow 0$ :

$$(2) \quad \begin{aligned} \frac{\partial N}{\partial t} &= r_+ N + r_0 E + r_- S + r_0 W - v \frac{\partial N}{\partial y}, \\ \frac{\partial S}{\partial t} &= r_- N + r_0 E + r_+ S + r_0 W + v \frac{\partial S}{\partial y}, \\ \frac{\partial E}{\partial t} &= r_0 N + r_+ E + r_0 S + r_- W - v \frac{\partial E}{\partial x}, \\ \frac{\partial W}{\partial t} &= r_0 N + r_- E + r_0 S + r_+ W + v \frac{\partial W}{\partial x}. \end{aligned}$$

In taking the limit  $\tau \rightarrow 0$ , we have introduced the following quantities:

$$(3) \quad \lim_{\tau \rightarrow 0} \frac{(p_+ - 1)}{\tau} = r_+,$$

$$(4) \quad \lim_{\tau \rightarrow 0} \frac{p_-}{\tau} = r_-,$$

$$(5) \quad \lim_{\tau \rightarrow 0} \frac{p_0}{\tau} = r_0,$$

$$(6) \quad \lim_{\tau \rightarrow 0} \frac{\lambda}{\tau} = v.$$

The set of equations (2) is formally soluble by transform methods, but the solutions are excessively complicated to be of use. The course we shall adopt here is to express the content of equations (2) in the form of a single equation in the total concentration of particles at a point. The decomposition into four directed "species" is unnecessary in any application that can be foreseen. The total particle concentration is the sum of the four "species" concentrations:

$$(7) \quad Q = N + S + E + W.$$

By a straightforward sequence of additions and subtractions of equations (2) followed by appropriate cross-differentiations, a single equation in  $Q$  may be obtained:

$$(8) \quad \left( \frac{\partial}{\partial t} + r \right)^2 \left( \frac{\partial}{\partial t} + 4r_0 \right) \frac{\partial Q}{\partial t} - v^2 \left( \frac{\partial}{\partial t} + r \right) \frac{\partial}{\partial t} \nabla^2 Q - 2r_0 v^2 \frac{\partial}{\partial t} \nabla^2 Q - 2r_0 v^2 \nabla^2 Q + v^4 \frac{\partial^4}{\partial x^2 \partial y^2} Q = 0.$$

We have now taken down the scaffold, but we have left the original anisotropy in the fourth-order derivative. This can be averaged over all co-ordinate systems obtained by rotations. The result is the replacement of

$$(9) \quad \frac{\partial^4 Q}{\partial x^2 \partial y^2} \text{ by } \frac{1}{8} \nabla^4 Q.$$

The final equation now reads:

$$(10) \quad \left( \frac{\partial}{\partial t} + r \right)^2 \left( \frac{\partial}{\partial t} + 4r_0 \right) \frac{\partial Q}{\partial t} - v^2 \left( \frac{\partial}{\partial t} + r \right) \frac{\partial}{\partial t} \nabla^2 Q - 2r_0 v^2 \frac{\partial}{\partial t} \nabla^2 Q - 2r_0 v^2 \nabla^2 Q + \frac{1}{8} v^4 \nabla^4 Q = 0$$

where there has been used the abbreviation

$$(11) \quad r = r_- - r_+.$$

If the approximation to equation (10) in which all time derivatives higher than the first and all space derivatives higher than the second are neglected there remains the classical form

$$(12) \quad \frac{\partial Q}{\partial t} = \frac{v^2}{2r} \nabla^2 Q.$$

If terms to second order in both space and time derivatives are retained, there remains

$$(13) \quad \frac{\partial Q}{\partial t} = \frac{v^2}{2r} \nabla^2 Q - \left( \frac{1}{4r_0} + \frac{2}{r} \right) \frac{\partial^2 Q}{\partial t^2}.$$

This equation is, in form, the same as the naive generalization of the one-dimensional Goldstein-Michelson equation obtained by averaging space derivatives over all directions as described before.

It is interesting to see what velocity autocorrelations are implied by the model we have utilized. This can be determined by calculating the second moments of the distribution of diffusate, i.e.,  $\langle x_i x_k \rangle$  defined by

$$(14) \quad \langle x_i x_k \rangle = \iint x_i x_k Q dx dy.$$

The velocity autocorrelations can then be found by the well-known kinematical theorem (see, for example, Pai 1957).

$$(15) \quad \frac{d^2}{dt^2} \langle x_i x_k \rangle = \overbrace{2v_i(t')v_k(t'+t)}^{\text{wavy bar}} = 2R_{ik}(t).$$

The wavy bar indicates a Lagrangian time average, and  $R_{ik}$  is the customary symbol for the velocity autocorrelation function. The calculation yields the following autocorrelation functions:

$$(16) \quad R_{ik}(t) = \delta_{ik} \frac{v^2}{4r_0 - r} [(2r_0 - r)e^{-rt} + 2r_0 e^{-4r_0 t}].$$

### III. THE THREE-DIMENSIONAL CASE

From a lattice diagram like that of Fig. (1) with additional planes above and below that pictured, it is a simple matter to write down a set of difference equations analogous to equations (1) which include the upward and downward moving species or components. The diffusing particle now has the possibility of turning upwards and downwards as well as to the right and to the left. In order to avoid unnecessary complication, we shall take these additional turning probabilities to be equal to the left-right turning probabilities designated, as before, by  $p_0$ .  $U^{i,j,k}$  and  $D^{i,j,k}$  represent the two new component species. The equations obtained in this way are:

$$(17) \quad \begin{aligned} N_{m+1}^{i,j+1,k} &= p_+ N_m^{i,j,k} + p_- S_m^{i,j,k} + p_0 E_m^{i,j,k} + p_0 W_m^{i,j,k} + p_0 U_m^{i,j,k} + p_0 D_m^{i,j,k}, \\ S_{m+1}^{i,j-1,k} &= p_- N_m^{i,j,k} + p_+ S_m^{i,j,k} + p_0 E_m^{i,j,k} + p_0 W_m^{i,j,k} + p_0 U_m^{i,j,k} + p_0 D_m^{i,j,k}, \\ E_{m+1}^{i+1,j,k} &= p_0 N_m^{i,j,k} + p_0 S_m^{i,j,k} + p_+ E_m^{i,j,k} + p_- W_m^{i,j,k} + p_0 U_m^{i,j,k} + p_0 D_m^{i,j,k}, \\ W_{m+1}^{i-1,j,k} &= p_0 N_m^{i,j,k} + p_0 S_m^{i,j,k} + p_- E_m^{i,j,k} + p_+ W_m^{i,j,k} + p_0 U_m^{i,j,k} + p_0 D_m^{i,j,k}, \\ U_{m+1}^{i,j,k+1} &= p_0 N_m^{i,j,k} + p_0 S_m^{i,j,k} + p_0 E_m^{i,j,k} + p_0 W_m^{i,j,k} + p_+ U_m^{i,j,k} + p_- D_m^{i,j,k}, \\ D_{m+1}^{i,j,k-1} &= p_0 N_m^{i,j,k} + p_0 S_m^{i,j,k} + p_0 E_m^{i,j,k} + p_0 W_m^{i,j,k} + p_- U_m^{i,j,k} + p_+ D_m^{i,j,k}. \end{aligned}$$

As before, we must pass to the limit  $\tau \rightarrow 0$  in order to form differential equations. Before doing so, we note the following relationship among the probabilities:

$$(18) \quad p_+ + p_- + 4p_0 = 1.$$

We introduce the definitions

$$\lim_{\tau \rightarrow 0} \frac{p_+ - 1}{\tau} = r_+,$$

$$(19) \quad \lim_{\tau \rightarrow 0} \frac{p_-}{\tau} = r_-,$$

$$\lim_{\tau \rightarrow 0} \frac{p_0}{\tau} = r_0,$$

and note that from (18), we have the identity

$$(20) \quad r_+ + r_- + 4r_0 = 0.$$

To simplify the writing of the differential equations obtained from (17), we make the additional definitions:

$$(21) \quad \begin{aligned} t_0 &= r_0 t, & \alpha &= r_+/r_0, \\ q_k &= x_k(r_0/v), & \beta &= r_-/r_0, \\ \dot{N} &= \partial N / \partial t, & N_k &= \partial N / \partial x_k, \end{aligned}$$

with

$$(22) \quad \alpha + \beta + 4 = 0$$

following from equation (20). The differential equations are, in the notations given above,

$$(23) \quad \begin{aligned} N_2 + \dot{N} &= \alpha N + \beta S + E + W + U + D, \\ -S_2 + \dot{S} &= \beta N + \alpha S + E + W + U + D, \\ E_1 + \dot{E} &= N + S + \alpha E + \beta W + U + D, \\ -W_1 + \dot{W} &= N + S + \beta E + \alpha W + U + D, \\ U_3 + \dot{U} &= N + S + E + W + \alpha U + \beta D, \\ -D_3 + \dot{D} &= N + S + E + W + \beta U + \alpha D. \end{aligned}$$

As before, a single differential equation may be obtained from this simultaneous set for the total concentration  $Q$ :

$$(24) \quad Q = N + S + E + W + U + D.$$

By a sequence of additions, subtractions, and cross-differentiations of the equations (23) this single equation may be obtained, and it takes the unwieldy form:

$$(25) \quad 2 \left( a - \frac{\partial}{\partial t_0} \right) \sum_{i=1}^3 \prod_{k \neq i}^3 \left( D^2 - \frac{\partial^2}{\partial q_k^2} \right) Q = \prod_{i=1}^3 \left( D^2 - \frac{\partial^2}{\partial q_i^2} \right) Q$$

where the following notations have been used:

$$(26) \quad D^2 = \left( \frac{\partial}{\partial t_0} - a \right) \left( \frac{\partial}{\partial t_0} - b \right),$$

$$(27) \quad a = \alpha - \beta = 4 + 2 \frac{r_+}{r_0} = - \left( 4 + 2 \frac{p_-}{p_0} \right),$$

$$(28) \quad b = \alpha + \beta - 2 = -6.$$

The last equalities hold by virtue of equation (22). If this differential equation is written out in full, there appear the following differential operators:

$$(29) \quad D_0 = \partial_1^2 + \partial_2^2 + \partial_3^2 = \nabla_q^2,$$

$$(30) \quad D_1 = \partial_1^4 + \partial_2^4 + \partial_3^4,$$

$$(31) \quad D_2 = \partial_1^2 \partial_2^2 + \partial_1^2 \partial_3^2 + \partial_2^2 \partial_3^2,$$

$$(32) \quad D_3 = \partial_1^2 \partial_2^2 \partial_3^2,$$

where we have written

$$(33) \quad \partial_i = \frac{\partial}{\partial q_i}$$

for brevity. In order to make these differential operators assume an isotropic form, it is necessary to average them over all possible rotations of the rectangular co-ordinate system, i.e., average them over the group of rotations in three dimensions. This results in the following replacements:

$$(34) \quad D_1 \rightarrow \frac{3}{20} \nabla_q^4,$$

$$(35) \quad D_2 \rightarrow \frac{1}{5} \nabla_q^4,$$

$$(36) \quad D_3 \rightarrow \frac{1}{105} \nabla_q^6.$$

Using these symmetrized operators in the expanded form of equation (25) yields

$$(37) \quad 2 \left( \frac{\partial}{\partial t_0} - a \right) \left[ 3D^4 - 2D^2 \nabla_q^2 + \frac{1}{5} \nabla_q^4 \right] Q = - \left[ D^6 - D^4 \nabla_q^2 + \frac{1}{5} D^2 \nabla_q^4 - \frac{1}{105} \nabla_q^6 \right] Q.$$

In the approximation in which all operators of higher than second order are neglected, this cumbersome equation reduces to the form of the telegraph equation in three dimensions. This approximate result has been obtained for the diffusion of heat, i.e., heat conduction, in gases by Catteneo (1958) and others.

The implied velocity autocorrelation functions can be calculated as before in the two-dimensional case. The result is:

$$(38) \quad R_{ik}(t) = \delta_{ik} v^2 \left[ \frac{a(a-10)}{(a+6)(a+12)} e^{ar_0 t} + \frac{4}{(a+6)} e^{-6r_0 t} + \frac{4}{(a+12)} e^{-12r_0 t} \right].$$

From the definition of the quantity  $a$  as given in equation (27), it can be seen that in the interesting case in which there is no backscattering in the diffusive encounters, one has

$$(39) \quad a = -4$$

and so equation (38) takes the especially simple form

$$(40) \quad R_{ik}(t) = \delta_{ik} \frac{v^2}{2} \left[ 7e^{-4r_0 t} + 4e^{-6r_0 t} + e^{-12r_0 t} \right].$$

## IV. DISCUSSION

The simple classical diffusion equation (equation (12)) is adequate for most applications in which the passage of the diffusive "wavefront" is not of interest. If this latter is of interest, the telegraph-type diffusion equation is a good description (equation (13)). Both of these equations can be obtained from simple one-dimensional theories extended to two and three dimensions simply by including derivatives along the additional co-ordinate axes. This is not, of course, a valid way of extending the one-dimensional results because it amounts simply to superposing one-dimensional diffusion in all directions. It neglects the possibility of turning motions altogether. It seems, however, that the complications produced by the more careful treatment given here are of limited practical value. The reasons for this are as follows: the equations obtained are of inconveniently high order, and they predict results significantly different from simpler equations (e.g., equations (12) or (13)), only when very high frequencies or transient effects are considered. Usually, however, these are of little importance and are, moreover, difficult to verify with any accuracy. The equations (16) and (38), which give the mean square spread of the diffusate as a function of time, do not differ very importantly from the expression involving a single exponential obtained from the simpler telegraph equation. All of these results, moreover, describe processes in which the velocity changes are Markovian, while the basic problem of turbulent diffusion theory is to describe the case in which the hereditary effects are more general than those of a Markov process, and spatial as well as temporal in extension. A partially successful attempt to describe these more general stochastic diffusion processes has been given by the author elsewhere (Bourret 1960), and additional studies are in preparation.

The author wishes to express his thanks to Dr. Maria Steinberg and Dr. Bela Lengyel for their help and comments.

## REFERENCES

BOURRET, R. 1960. *Can. J. Phys.* **38**, 665, 1213.  
CATTENEO, C. 1958. *Compt. rend.* **247**, 431.  
DAVIES, R. W., DIAMOND, R. J., and SMITH, T. B. 1954. American Institute of Aerological Research Report, OSR-TN-54-62.  
GOLDSTEIN, S. 1951. *Quart. J. Mech. and Appl. Math.* **4** (2), 129.  
GUPTA, H. C. 1959. *J. Math. and Phys.* **38** (1), 36.  
MICHELSON, I. 1954. American Institute of Aerological Research Report, Contract AF 18(600)-476.  
PAI, SHIH-I. 1957. *Viscous flow theory, II, turbulent flow* (D. Van Nostrand Company, Inc., New York).  
TAYLOR, G. I. 1921. *Proc. London Math. Soc.* **20**, 196.

# MAXWELL'S EQUATIONS AND MATRIX ELEMENTS IN QUANTUM ELECTRODYNAMICS<sup>1</sup>

H. A. VENABLES

## ABSTRACT

Matrix elements of second-order processes in quantum electrodynamics are obtainable directly from the use of Maxwell's and Dirac's equations.

## INTRODUCTION

Several spinor formulations of the electromagnetic field have been given (Molière 1949; Archibald 1955; Good 1957; Moses 1959). When it is demanded that the spinor formulation lead to a spin of 1 in the same way as the electron theory of Dirac (1958) leads to a spin of  $\frac{1}{2}$  then the choice of formulation is limited.

In the case of the interaction between the electron and the electromagnetic field, a theory need only provide a spin of  $\frac{1}{2}$ . By considering the interaction exclusively a suitable spinor formulation can be developed for obtaining second-order matrix elements in a manner which appears to have been overlooked hitherto. The technique is of interest because the Hamiltonian and Poynting vector densities as well as a generalized electromagnetic force enter the theory automatically.

## SPINOR FORMULATION

Maxwell's equations can be written as

$$(1) \quad \left( \frac{\partial}{\partial t} \pm \mathbf{d} \cdot \frac{\partial}{\partial \mathbf{x}} \right) \mathbf{d} \cdot (\mathbf{E} \pm i\mathbf{H}) = 4\pi(\pm \rho(x) - \mathbf{d} \cdot \mathbf{j}(x))$$

where  $\mathbf{d}$  denotes Pauli's matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and  $c = 1$ .

In order to obtain the equation in conventional form, multiply out the left-hand side, collect the terms together which are associated with the same resulting matrix, and then separate real and imaginary parts for the required equations.

The two equations of (1) with upper and lower signs may be put into one equation

$$\left( \begin{pmatrix} \frac{\partial}{\partial t} + \mathbf{d} \cdot \frac{\partial}{\partial \mathbf{x}} \\ \frac{\partial}{\partial t} - \mathbf{d} \cdot \frac{\partial}{\partial \mathbf{x}} \end{pmatrix} \begin{pmatrix} -\mathbf{d} \cdot (\mathbf{E} - i\mathbf{H}) \\ \mathbf{d} \cdot (\mathbf{E} + i\mathbf{H}) \end{pmatrix} \right) = 4\pi \begin{pmatrix} \rho(x) - \mathbf{d} \cdot \mathbf{j}(x) \\ \rho(x) + \mathbf{d} \cdot \mathbf{j}(x) \end{pmatrix}$$

<sup>1</sup>Manuscript received July 27, 1960.

Contribution from the University of Alberta, Edmonton, Alberta.

This will be written more briefly as

$$(2) \quad \begin{aligned} \gamma \frac{\partial}{\partial x} \phi(x) &= 4\pi\gamma j(x), \\ x &= \mathbf{x}, t, \\ j &= \mathbf{j}, \rho, \end{aligned}$$

where the  $\gamma$  matrices are

$$\gamma = \begin{pmatrix} 0 & \mathbf{0} \\ -\mathbf{0} & 0 \end{pmatrix} \quad \gamma_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Equation (1) has a Hermitean conjugate with differentiation acting to the left

$$(3) \quad \phi(x) \gamma \frac{\partial}{\partial x} = -4\pi\gamma j(x).$$

The Hamiltonian and Poynting vector densities occur in

$$(4) \quad \phi^+ \phi = (E^2 + H^2) + 2\gamma \gamma \mathbf{E} \times \mathbf{H}$$

and the electromagnetic power and force is the real part of

$$(5) \quad \text{Re}(\gamma j \phi) = -(\mathbf{E} \cdot \mathbf{j}) + \gamma \gamma (\rho \mathbf{E} + \mathbf{j} \times \mathbf{H}).$$

Equation (2) has a form similar to Dirac's equation, which will be written as

$$(6) \quad \left( \gamma \frac{\partial}{\partial x} + im \right) \psi(x) = 0.$$

If we premultiply (6) by  $\phi(x)$  and postmultiply (3) by  $\psi(x)$  and then add we obtain the equation of interaction

$$(7) \quad \phi(x) \left( \gamma \frac{\partial}{\partial x} + im \right) \psi(x) = 0.$$

The most general solution of this equation would be difficult to find. However, it will be shown that by choosing a reasonable Green's function matrix elements are obtainable, agreeing with those given by more elaborate methods (for example, Akhiezer and Berestetsky 1953; Källen 1958). Equation (7) leads to the desired spin,  $\frac{1}{2}$ , for the interaction between the two fields.

#### MATRIX ELEMENTS

Let the integral form corresponding to (7) be

$$(8) \quad \phi(x) \psi(x) = \phi_0(x) \psi_0(x) + 4\pi \int G(x, x') \gamma j(x') \psi(x') d^4 x'$$

where  $G(x, x')$  is an appropriate Green's function.

Multiply equation (8) by its Hermitean conjugate to give, where a minor term is omitted,

$$\begin{aligned} & \psi^+(x) \phi^+(x) \phi(x) \psi(x) - \psi_0^+(x) \phi_0^+(x) \phi_0(x) \psi_0(x) \\ &= 4\pi \{ \int \psi^+(x') \gamma^+ j(x') G^+(x', x) d^4 x' \phi_0(x) \psi_0(x) + \psi_0^+(x) \phi_0^+(x) \int G(x, x') \\ & \quad \gamma j(x') \psi(x') d^4 x' \}. \end{aligned}$$

The two terms on the left-hand side are seen from (4) to be probability densities of the Hamiltonian and Poynting vector densities. The difference of these two terms can be regarded as the change in these densities due to scattering. The terms on the right-hand side with the factors  $\phi \gamma j$  are seen from (5) to be generalized forces. Figure 1 shows the corresponding scattering diagrams.



FIG. 1.

The matrix elements of second-order processes consist of two terms. The first term is derived and then by analogy a second term is fitted to the second diagram of Fig. 1. Here such an argument is not required since two terms appear. With a minor modification to the second integral we have

$$4\pi \int \psi^+(x') \{ \gamma^+ j(x') G^+(x', x) \phi_0(x) + \phi_0^+(x) G(x, x') \gamma j(x') \} \psi_0(x) d^4 x d^4 x'.$$

Matrix elements are usually evaluated in the momentum representation, for which the following substitutions are required,

$$\begin{aligned} \psi(x) &= (1/2\pi)^{3/2} (m/2E)^{1/2} u(p_1) e^{ip_1 x} \\ \psi^+(x) &= (1/2\pi)^{3/2} (m/2E)^{1/2} u^+(p_2) e^{-ip_2 x} \\ j(x') \phi(x) &= A j(k_2) \phi(k_1) e^{i(k_1 x - k_2 x')} \\ \phi^+(x) j(x') &= A \phi^+(k_2) j(k_1) e^{-i(k_2 x - k_1 x')} \\ G(x, x') &= \frac{-2i}{(2\pi)^4} \int \frac{e^{ip(x-x')}}{(\gamma p - m)} d^4 p \\ px &= \mathbf{p} \cdot \mathbf{x} - Et. \end{aligned}$$

The normalization constant  $A$  will depend on the number of absorbed and emitted photons. For one such photon as in Bremsstrahlung

$$A = e^2 (\frac{1}{2}\pi)^{3/2} (\hbar c/2)^{1/2} (1/\omega)^{1/2}.$$

For two such photons as in Compton scattering

$$A = e^2 (\frac{1}{2}\pi)^{3/2} (\hbar c/2) (1/\omega \omega')^{1/2}.$$

After substitution and integration the following is obtained

$$u^+(p_2) \left\{ \gamma j(k_2) \frac{i\gamma f_1 - m}{f_1^2 + m^2} \phi(k_1) + \phi^+(k_2) \frac{i\gamma f_2 - m}{f_2^2 + m^2} \gamma j(k_1) \right\} u(p_1) \delta(p_1 - p_2 + k_1 - k_2)$$

$$f_1 = (p_1 + k_1), \quad f_2 = (p_2 - k_1).$$

If only  $\mathbf{j}$  and  $\mathbf{E}$ , the factors of electromagnetic power  $\mathbf{E} \cdot \mathbf{j}$ , are considered then we obtain

$$u^+(p_2) \left\{ \gamma \cdot \mathbf{j}(k_2) \frac{i\gamma f_1 - m}{f_1^2 + m^2} \gamma \cdot \mathbf{E}(k_1) + \gamma_4 \gamma \cdot \mathbf{E}(k_2) \frac{i\gamma f_2 - m}{f_2^2 + m^2} \gamma \cdot \mathbf{j}(k_1) \right\} u(p_1) \delta(p_1 - p_2 + k_1 - k_2).$$

This can be compared with the expression for the Compton scattering in which the four potentials appear in place of the current and the electric field vector (Akhiezer and Berestetsky 1953; Källen 1958).

If the Lorentz force factors are considered, then the matrix element of Bremsstrahlung is found. Also in a similar manner by using positron wave functions the matrix elements of pair production and two-photon annihilation are derived.

The author would like to thank the International Nickel Company of Canada for a fellowship.

#### REFERENCES

AKHIEZER, A. I. and BERESTETSKY, V. B. 1953. Quantum Electrodynamics (English translation AEC-tr-2876, Washington D.C.).  
 ARCHIBALD, W. J. 1955. Can. J. Phys. **33**, 565.  
 DIRAC, P. A. M. 1958. Principles of quantum mechanics (Oxford University Press, London).  
 GOOD, R. H., JR. 1957. Phys. Rev. **105**, 1914.  
 KÄLLEN, A. O. G. 1958. Handbuch der Physik Band V/I (Springer-Verlag, Berlin).  
 MOLIÈRE, G. 1949. Ann. Physik, **6**, 146.  
 MOSES, H. E. 1959. Phys. Rev. **113**, 1670.

# AN ELECTRON SPIN RESONANCE STUDY OF MANGANESE IMPURITY IN BRUCITE<sup>1</sup>

W. A. PIECZONKA,<sup>2</sup> H. E. PETCH, AND A. B. McLAY

## ABSTRACT

A single crystal of brucite,  $Mg(OH)_2$ , containing manganese impurity to the extent of 100 parts per million, has been studied by means of the electron spin resonance (esr) technique. The study was made at room and liquid air temperatures. The observed absorption spectrum has been successfully interpreted in terms of parameters found in the appropriate spin-Hamiltonian. The measured values of these parameters at room temperature were found to be:

$$\begin{array}{ll} g_{\parallel} = 2.0001 \pm 0.0005, & g_{\perp} = 2.0005 \pm 0.0005, \\ D = -7.20 \pm 0.25 \times 10^{-4} \text{ cm}^{-1}, & a = +10.82 \pm 0.45 \times 10^{-4} \text{ cm}^{-1}, \\ A = -85.7 \pm 0.4 \times 10^{-4} \text{ cm}^{-1}, & B = -84.9 \pm 0.6 \times 10^{-4} \text{ cm}^{-1}. \end{array}$$

## INTRODUCTION

When paramagnetic ions are located in a solid, they are influenced by the magnetic and electric fields produced by their surroundings. Certain aspects of the interaction between a paramagnetic ion and its diamagnetic neighbors may be studied using the electron spin resonance (esr) technique, provided the interaction between the paramagnetic ion and its paramagnetic neighbors, which broadens the resonance signals, is kept very small. This is accomplished by working with very low concentrations of paramagnetic ions.

The extent to which the energy levels of the paramagnetic ion are affected by the crystalline electric field depends upon the species of ion in question and upon the strength and symmetry of the electric field. Most esr studies involve highly symmetrical atomic groupings which produce fields of spherical or near spherical symmetry. A spherically symmetric field is referred to as a cubic field. A field that departs from spherical symmetry through a slight distortion along one axis can be considered to be composed of, primarily, a cubic field plus a small axial field component. The strength of the electric field interaction is important in that it may very strongly affect the magnetic behavior of the ions of the various transition groups. For example, the  $^4F$  ground state of a free  $Cr^{3+}$  ion is split by an electric field interaction into three widely separated levels with an orbital singlet lying lowest. Thus, the effective magnetic moment of the ion in the field is reduced to a "spin-only" value. It is customary to express the strength of the electric field interaction relative to some other interaction within the free ion. In the case of manganese, as well as the other  $3d$  transition elements, it is moderate, i.e., it is larger than the spin-orbit but smaller than the Coulomb interaction.

<sup>1</sup>Manuscript received September 13, 1960.

Contribution from the Departments of Physics and Metallurgy, Hamilton College, McMaster University, Hamilton, Ontario

<sup>2</sup>Holder of a Shell Oil Company Scholarship (1958-59) and a Cominco Fellowship (1959-60) awarded by the Shell Oil Company of Canada and the Consolidated Mining and Smelting Company of Canada, respectively.

Rabi (1927) provided the first experimental evidence of *S*-state splitting in an electric field when he observed small anisotropies in the magnetic susceptibility of manganese salts. Soon after, Bethe (1929) predicted from group theoretical considerations that the ground state of an *S*-state ion would be split by an electric field even if the field possessed cubic symmetry. Later, Van Vleck and Penney (1934) suggested that a mechanism involving spin-orbit coupling with excited states might be responsible for this splitting. With the advent of esr it became possible to study this effect with greater precision than could be achieved with susceptibility measurements.

The first detailed esr experiment involving  $Mn^{2+}$  was performed by Bleaney and Ingram (1951). They studied the manganese ion in a nearly cubic electric field and found that the ground state was split into doublets corresponding to different  $\pm M$  values, where  $M$  is the magnetic spin quantum number. The observed ratio of the so-called axial to cubic field splitting was approximately 15:1, which is much larger than could be accounted for in terms of the mechanism suggested by Van Vleck and Penney. Thus it became apparent that an additional mechanism was operative in the case of axial field splitting. Abragam and Pryce (1951) qualitatively interpreted this axial field splitting in terms of an axially distorted spherical electron cloud distribution and spin-spin interaction within the ion itself. In theory, this interpretation postulates admixtures to the *S*-state of electron configurations which yield resultant non-zero orbital angular momenta. Qualitative approaches to this problem have not been entirely successful. Because the  $Mn^{2+}$  ion is an *S*-state ion, it has been studied by esr techniques to a greater extent than any other paramagnetic ion.

It is clear that the basic mechanisms responsible for the *S*-state splitting are not completely understood, and more work is required. Much of the early work on manganese dealt with cases where the axial field effects were greater than those of the cubic field. Recently, Low (1957) has embarked on a program of investigating the  $Mn^{2+}$  ion in cubic fields in an effort to study only the cubic field contribution to the ground state splitting. In the present case we have studied the ion in a field of axial symmetry such that the two effects, axial and cubic, happen to be of the same order of magnitude.

#### THEORY

The spin-Hamiltonian generally used to describe an *S*-state ion situated in an electric field of axial symmetry and a uniform magnetic field is one given by Pryce (1950a). For  $S = 5/2$  the spin-Hamiltonian is written

$$(1) \quad \mathfrak{H} = \beta[g_{\parallel}H_zS_z + g_{\perp}(H_xS_x + H_yS_y)] + D\left(S_z^2 - \frac{35}{12}\right) + \frac{a}{6}\left(S_u^4 + S_v^4 + S_w^4 - \frac{707}{16}\right) + AI_zS_z + B(I_xS_x + I_yS_y)$$

where  $\beta$  is the Bohr magneton,

$g_{\parallel}$  is the component of the  $g$  tensor parallel to the symmetry axis,

$g_{\perp}$  is the component of the  $g$  tensor perpendicular to the symmetry axis,

*D* is the axial field splitting parameter,

*a* is the cubic field splitting parameter,

*A* is the component of the hyperfine splitting parameter parallel to the symmetry axis,

*B* is the component of the hyperfine splitting parameter perpendicular to the symmetry axis,

*x,y,z* is an orthogonal, crystal co-ordinate system,

*u,v,w* is an orthogonal, magnetic complex co-ordinate system.

For trigonal crystal symmetry the co-ordinate systems are chosen so that the symmetry axis, *z*, is the (111) direction in the *u,v,w* system.

From the spin-Hamiltonian the six electronic energy levels corresponding to *S* = 5/2 can be derived. Allowed transitions between these levels according to the selection rule *M* = ±1 result in a spectrum of five so-called "fine lines". In addition, because the manganese nucleus also has a spin, *I* = 5/2, each fine line is split into six hyperfine lines. Thus a complete Mn<sup>2+</sup> spectrum of 30 lines results for each unique ionic site. It can be shown, Bleaney (1951), that for axial symmetry the resonant fields for the fine line transitions can be expressed as follows:

$$(2) \quad H_{M \leftrightarrow M-1} = H_0 - D(M - \frac{1}{2}) \left( 3 \frac{g_{\parallel}^2}{g^2} \cos^2 \theta - 1 \right) + \frac{D^2}{2H_0} \left[ \frac{g_{\parallel} g_{\perp}}{g^2} \right]^2 [4S(S+1) - 24M(M-1) - 9] \cos^2 \theta \sin^2 \theta - \frac{D^2}{8H_0} \left( \frac{g_{\perp}}{g} \right)^4 [2S(S+1) - 6M(M-1) - 3] \sin^4 \theta + f(a)$$

where  $\theta$  is the angle between the symmetry axis and the applied magnetic field, *f(a)* is a function of crystal symmetry, and *H*<sub>0</sub> and *g* are respectively defined by  $g\beta H_0 = \hbar\nu$  and  $g^2 = g_{\parallel}^2 \cos^2 \theta + g_{\perp}^2 \sin^2 \theta$ . Each fine line splits into hyperfine lines whose field positions, relative to their parent fine line, are given by

$$(3) \quad - \left\{ Km + \frac{B^2}{4H_0} \left[ \frac{A^2 + K^2}{K^2} \right] [I(I+1) - m^2] + \frac{B^2}{2H_0} \left( \frac{A}{K} \right) (2M-1)m + \frac{1}{2H_0} \left[ \frac{A^2 - B^2}{K} \right]^2 \left[ \frac{g_{\parallel} g_{\perp}}{g^2} \right]^2 m^2 \sin^2 \theta \cos^2 \theta \right\}$$

where *m* is the nuclear magnetic spin quantum number and *K* is defined by  $K^2 g^2 = A^2 g_{\parallel}^2 \cos^2 \theta + B^2 g_{\perp}^2 \sin^2 \theta$ . Since for the Mn<sup>2+</sup> ion  $g_{\parallel} \approx g_{\perp}$ , the fine line positions for the case  $H_0 \gg D$  can be written,

$$(4) \quad H_{\pm 5/2 \leftrightarrow \pm 3/2} = H_0 \mp 2D(3 \cos^2 \theta - 1) - \frac{32D^2}{H_0} \cos^2 \theta \sin^2 \theta + \frac{D^2}{H_0} \sin^4 \theta \mp 2Pa,$$

$$H_{\pm 3/2 \leftrightarrow \pm 1/2} = H_0 \mp D(3 \cos^2 \theta - 1) + \frac{4D^2}{H_0} \cos^2 \theta \sin^2 \theta - \frac{5D^2}{4H_0} \sin^4 \theta \pm \frac{5}{2}Pa,$$

$$H_{+1/2 \leftrightarrow -1/2} = H_0 + \frac{16D^2}{H_0} \cos^2 \theta \sin^2 \theta - \frac{2D^2}{H_0} \sin^4 \theta,$$

where  $P = 1 - 5\phi$  and  $\phi = p^2q^2 + q^2r^2 + r^2p^2$  and  $p, q, r$  are direction cosines of the applied field direction referred to the axes  $u, v, w$ . Also, since  $A = B$  and  $H_0 > A$ , expression (3) can be well approximated by

$$(5) \quad - \left[ Km + \frac{B^2}{4H_0} \left( \frac{A^2 + K^2}{K^2} \right) \left( \frac{35}{4} - m^2 \right) + \frac{B^2}{2H_0} \left( \frac{A}{K} \right) (2M-1)m \right].$$

The magnitudes of the various parameters in the spin-Hamiltonian can be easily obtained upon equating the measured line positions to those given by (4) and (5). The signs of the field parameters can be obtained by observing the spectrum at low temperatures since by virtue of Boltzmann distributions, the intensity of the  $-3/2 \leftrightarrow -5/2$  transition will be greater than that of the  $+5/2 \leftrightarrow +3/2$  transition. This feature can be used to identify the lines and to assign correct signs to  $D$  and  $a$ . One must, however, be careful with regard to the possible occurrence of a phase transition since it is conceivable that the sign of  $D$  could change as the crystal is cooled through the transition temperature. The signs of  $A$  and  $B$  can be determined from second-order effects in the hyperfine structure. From (5) it can be shown that for  $A$  positive, transitions involving positive  $m$  occur at lower fields than transitions involving negative  $m$ . The converse applies for  $A$  negative. Now the separation at  $\theta = 0$  between the outermost hyperfine lines arising from the  $+5/2 \leftrightarrow +3/2$  transition are  $5A + 10B^2/H_0$  for positive  $A$  and  $-5A - 10B^2/H_0$  for negative  $A$ . Similarly, the separations between the outermost hyperfine lines arising from the  $-3/2 \leftrightarrow -5/2$  transition are  $5A - 10B^2/H_0$  for positive  $A$  and  $-5A + 10B^2/H_0$  for negative  $A$ . Provided that the fine lines can be identified, the signs of  $A$  and  $B$  can easily be determined because the second-order correction,  $10B^2/H_0$ , for  $Mn^{2+}$  is of the order of 25 gauss when the experiment is performed at X-band frequencies.

#### THE BRUCITE CRYSTAL

The crystal structure of brucite,  $Mg(OH)_2$ , was first determined by Aminoff (1919) and was shown to be of the  $CdI_2$  type with space group  $P\bar{3}m1$ . The unit cell is trigonal and contains one formula unit. The magnesium ions lie in the  $(0001)$  planes, while the oxygen ions lie in planes  $0.22c$  above and below the  $(0001)$  planes. The hydrogen atom associated with each oxygen ion lies on a line through the oxygen parallel to the  $c$ -axis and at a distance of  $0.98 \text{ \AA}$  removed from it (Elleman and Williams 1956). As can be seen from Fig. 1, the magnesium ion is surrounded by six oxygen ions situated at the corners of a near perfect octahedron. With a manganese ion in place of a magnesium ion this octahedron constitutes the so-called magnetic complex and, in usual esr terminology, is designated as  $[\text{Mn}(\text{OH})_6]^{4-}$ . The magnetic complex is distorted from octahedral symmetry through a slight compression along the crystallographic  $c$ -axis, which is the  $(111)$  direction in the  $u, v, w$  co-ordinate system.

Brucite crystals are usually found as thin, colorless plates: a variation called nemalite or broom brucite is also found, but is not as common. The  $c$ -axis lies in a direction perpendicular to the plates. The crystal has good

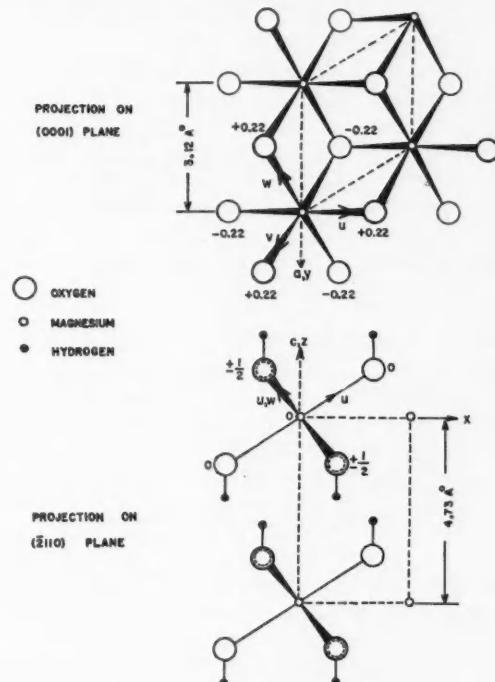


FIG. 1. Brucite crystal structure.

cleavage perpendicular to the *c*-axis, but the cleaved sections are soft and brittle.

Samples of brucite from three geographical locations—New Mexico, Pennsylvania, and Quebec—were investigated. The Quebec sample, containing approximately 100 parts per million manganese impurity as determined by spectrographic analysis, exhibited an absorption spectrum in which the individual lines were best resolved. The other samples exhibited a less resolved spectrum, perhaps because they were not as perfect and(or) contained a higher concentration of manganese.

Really good single crystals of brucite are difficult to obtain because a substructure, in which the small domains have their individual *c* axes oriented over a small angular range about the morphological *c*-axis, is very common. X-Ray analysis showed that the Quebec sample was representative of a reasonably good single crystal and this crystal was used throughout this investigation. The size of the crystal was approximately  $4 \times 4 \times 1$  mm.

#### APPARATUS AND EXPERIMENTAL PROCEDURE

The X-band spectrometer used in this study was constructed in this laboratory. As can be seen from the spectrometer block diagram, Fig. 2, it is of the

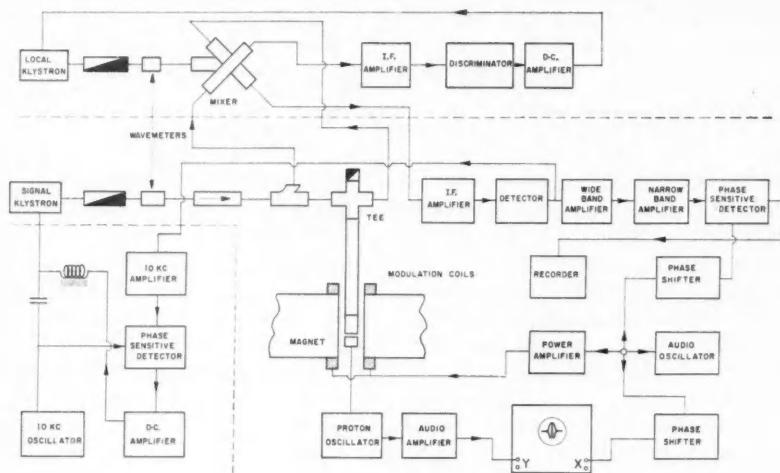


FIG. 2. Block diagram of esr spectrometer.

superheterodyne type. The actual spectrometer design was modelled to a certain extent on one described by Hirshon and Fraenkel (1955); it does differ, however, in the type of signal frequency control system used. Varian X-13B and V-290 klystrons are used as signal and local oscillators respectively. The theoretical minimum number of manganese ions that could be detected under ideal conditions in a brucite crystal with such a detection system is approximately  $3 \times 10^{11}$ . The actual number that could be detected in brucite with a signal-to-noise ratio of 2:1 was  $10^{13}$ .

As is customary in esr work, the applied magnetic field, produced by a Varian 6-in. model V-4007 electromagnet, was swept through the absorption spectrum at a desired rate and the frequency of the incident radiation was held constant. The magnetic field was calibrated at 25-gauss intervals by means of a proton resonance probe. An absolute calibration was obtained by relating the proton calibration points to the field at which a picryl hydrazyl absorption occurred. Over-all field calibration was accurate to within  $\pm 1.0$  gauss.

Most of the experimental work was done at room temperature. Some work, however, was done at liquid air temperature to determine the dependence of the spin-Hamiltonian parameters on temperature. Also, spectra were observed at  $67^\circ$  K in an effort to identify the absorption lines. This temperature was reached by pumping on the volume over a liquid air surface.

In the room temperature experiment the sample was mounted on a lucite rod and held at the center of a rectangular cavity which resonated at 9128 Mc/sec in the  $TE_{102}$  mode. The crystal was rotated about three axes which were, in turn, maintained perpendicular to the applied magnetic field, and

spectra were recorded at 10° intervals. These rotation axes were the *a*-axis, the *c*-axis, and an axis inclined at 30° and 90°, respectively, to the *a* and *c* axes. In some regions where the absorption line positions varied rapidly with the angle of rotation, spectra were also recorded at 5° intervals.

For the low temperature experiments the sample was glued to an *H*-plane wall, a quarter wavelength from the bottom, of a cavity whose resonant frequency in the TE<sub>108</sub> mode was 8567 Mc/sec at 90° K. The cavity was made long so that by immersing almost its complete length in liquid air, the sample temperature could be held constant. The sample could be oriented by rotating the whole cavity. Only the rotation about an *a*-axis was performed at 90° K.

#### EXPERIMENTAL RESULTS AND ANALYSIS

Typical room temperature spectra obtained with the *a*-axis perpendicular to  $\mathbf{H}$  and  $\theta = 0^\circ$  and  $90^\circ$  are shown in Figs. 3a and 3b respectively. The narrow line at the center of each spectrum is a hydrazyl marker. Interpretation of each spectrum is also given on the figures. Here, the hyperfine lines belonging to each fine line are drawn to the theoretical intensity scale of 5:8:9:8:5. Quantum numbers were assigned to the lines on the basis of changes in the integrated intensities of the outermost lines at  $\theta = 0$ ,  $T = 67^\circ$  K, and second-order effects in the hyperfine structure. Although at 67° K the theoretical intensity ratio of the  $-3/2 \leftrightarrow -5/2$  to the  $+5/2 \leftrightarrow +3/2$  transition is only 1.05, it was possible to make positive identification after several scans through the spectrum.

The magnetic field intensities at which the various transitions at room temperature for  $\theta = 0^\circ$  and  $\theta = 90^\circ$  occur are given in Table I. From expressions (4) and (5) it can be shown that certain relationships exist between the "centers of gravity" of the two central hyperfine lines and the fine lines to which they belong. These relationships were used to calculate the fine line positions. The points on Figs. 4a and 4b show the angular dependence of the fine line positions for rotations of the crystal at room temperature about the *a*-axis and the axis at 30° to the *a*-axis, respectively. The former dependence repeats after 180° and the latter after 90°.

Results for rotation about the *c*-axis are not given because the line positions should not change. In the actual experiment, very slight shifts in the line positions were observed. These were attributed to error in crystal alignment which was accurate only to within  $\pm 1^\circ$ . Figure 5 shows the angular dependence of the fine line positions for rotation of the crystal, maintained at 90° K, about an *a*-axis.

The following methods were used to determine the various spin-Hamiltonian parameters. The *g*-values,  $g_{\parallel}$ , and  $g_{\perp}$ , were obtained from values of  $H_0$  at  $\theta = 0^\circ$  and  $\theta = 90^\circ$ . At either orientation an average value of  $H_0$  was obtained as follows. From the positions of the hyperfine lines belonging to the central fine line the "centers of gravity" of each pair of hyperfine lines, designated by  $\pm m$ , were found. Relationships between these fields and  $H_0$  were used to calculate values of  $H_0$ . From the average of these  $H_0$  values the *g*-values

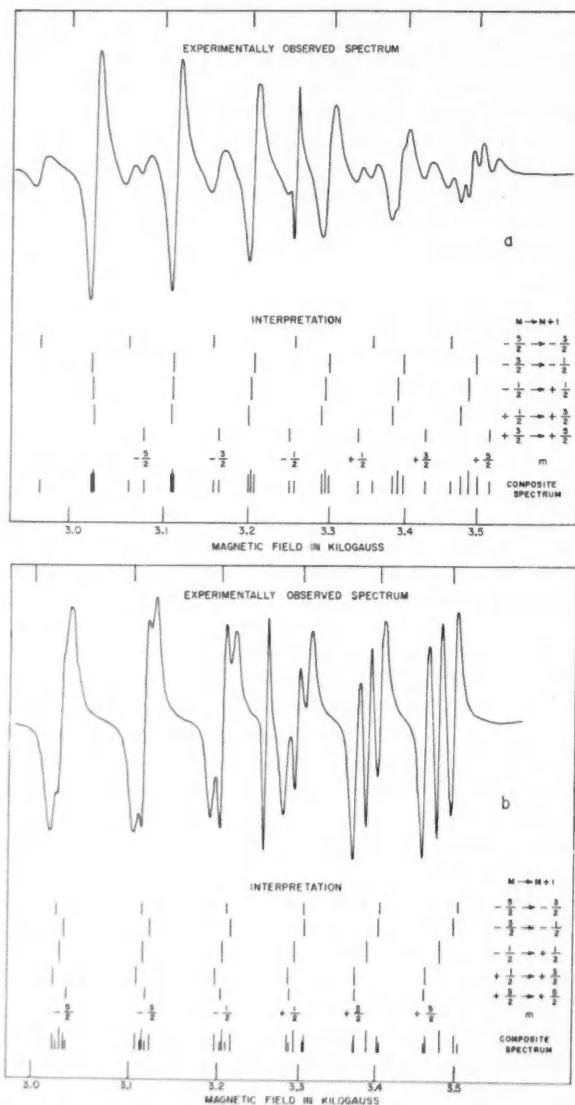


FIG. 3. Recordings of the derivative of the ESR absorption lines in brucite containing 100 parts per million manganese impurity. Recordings were made at room temperature and constant frequency of 9128 Mc/sec. (a) Applied magnetic field along *c*-axis. (b) Applied magnetic field at 90° to *c*-axis.

TABLE I

Position of esr absorption lines in gauss for  $\text{Mg}(\text{OH})_2:\text{Mn}$  at room temperature and frequency 9128 Mc/sec. Line positions accurate to  $\pm 1.5$  gauss

Transition	<i>m</i>	$\theta = 0$	Fine line position	$\theta = 90$	Fine line position
$-5/2 \leftrightarrow -3/2$	$-5/2$	2973.8		u	
	$-3/2$	3064.7		u	
	$-1/2$	3158.7*		u	
	$+1/2$	3255.7*	3217.8	u	
	$+3/2$	3356.0		u	
	$+5/2$	3460.6		u	
$-3/2 \leftrightarrow -1/2$	$-5/2$	u		u	
	$-3/2$	u		u	
	$-1/2$	3206.4*		3214.8*	
	$+1/2$	3300.5*	3254.0	3307.1*	3271.5
	$+3/2$	3397.1*		3402.1	
	$+5/2$	3496.2		3501.1	
$-1/2 \leftrightarrow +1/2$	$-5/2$	3027.7		3031.1	
	$-3/2$	3114.7		3116.0	
	$-1/2$	3204.4		3203.8	
	$+1/2$	3295.8	3260.7	3294.0	3259.5
	$+3/2$	3391.3		3386.5	
	$+5/2$	3485.5		3481.5	
$+1/2 \leftrightarrow +3/2$	$-5/2$	u		u	
	$-3/2$	u		u	
	$-1/2$	3202.4*		3195.0*	
	$+1/2$	3291.5*	3257.5	3283.0*	3247.6
	$+3/2$	3383.1*		3372.1	
	$+5/2$	3477.2		3461.6	
$+3/2 \leftrightarrow +5/2$	$-5/2$	3080.2		u	
	$-3/2$	u		u	
	$-1/2$	3252.0		u	
	$+1/2$	3336.5	3303.8	u	
	$+3/2$	3425.0		u	
	$+5/2$	3514.7		u	

NOTE: Line positions marked (\*) are accurate to  $\pm 3.0$  gauss.

Line positions marked (u) are unresolved.

were calculated. From expression (5) it can be seen that the separation between the outermost hyperfine lines belonging to the central fine line is  $K$ . From several values of  $K$ , several values of  $A$  and  $B$  were determined and averaged. The negative signs of these parameters were determined from the observation that the hyperfine separations belonging to the  $-3/2 \leftrightarrow -5/2$  transition are larger than those belonging to the  $+5/2 \leftrightarrow +3/2$  transition. Preliminary values of the field parameters  $D$  and  $a$  were obtained from expression (4) when these were equated to the actual fine line positions at  $\theta = 0^\circ$ . The value of  $D$  thus obtained was substituted for  $D$  in terms of (4) that contain  $D^2$ . These modified expressions were then used to solve for  $D$  and  $a$  at several values of  $\theta$ . The average values of these, along with the other parameters, are given in Table II.

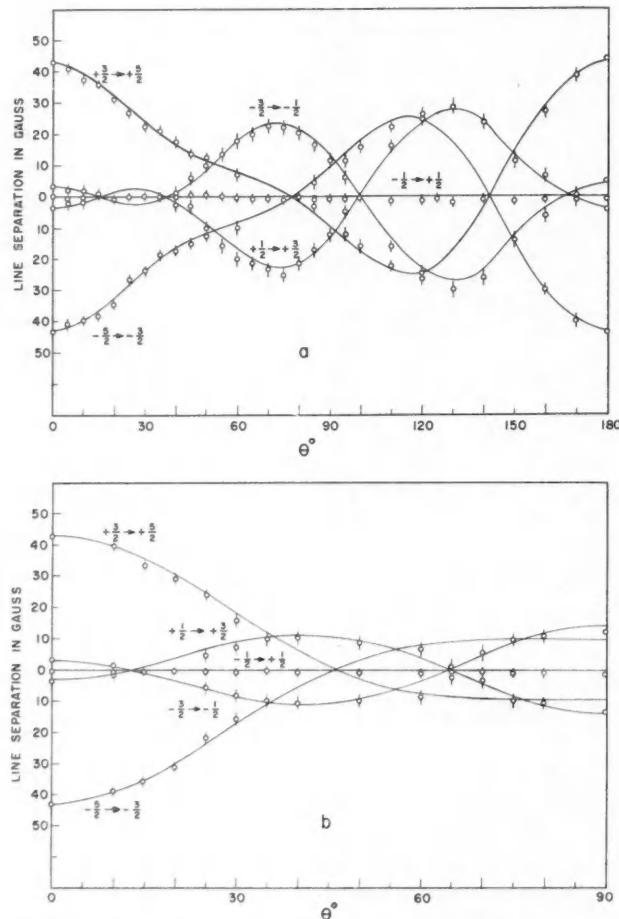


FIG. 4. Angular dependence of esr "fine line" positions for brucite at room temperature.  
 (a) Crystal rotated about  $a$ -axis. (b) Crystal rotated about an axis making inclinations of 30° and 90° respectively with the  $a$  and  $c$  axes.

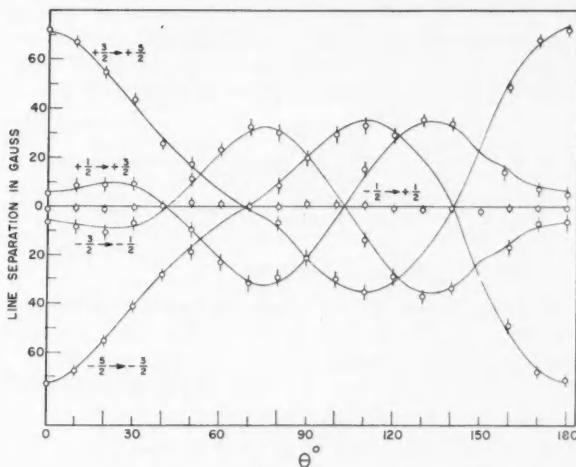


FIG. 5. Angular dependence of esr "fine line" positions for brucite at 90° K. Crystal rotated about *a*-axis.

TABLE II  
Measured values of the spin-Hamiltonian parameters for Mn<sup>2+</sup> in brucite

Parameter	295° K		90° K	
	Gauss	10 <sup>-4</sup> cm <sup>-1</sup>	Gauss	10 <sup>-4</sup> cm <sup>-1</sup>
<i>D</i>	-7.20 ± 0.25	-6.77 ± 0.25	-13.7 ± 0.25	-12.8 ± 0.25
<i>a</i>	+10.82 ± 0.45	+10.15 ± 0.45	+12.8 ± 0.25	+12.0 ± 0.25
<i>A</i>	-91.4 ± 0.4	-85.7 ± 0.4	-91.6 ± 1.0	-85.9 ± 1.0
<i>B</i>	-90.6 ± 0.6	-84.9 ± 0.6	-90.6 ± 0.6	-84.9 ± 0.6
<i>g</i> <sub>  </sub>	2.0001 ± 0.0005		2.0007 ± 0.0010	
<i>g</i> <sub>⊥</sub>	2.0005 ± 0.0005		2.0020 ± 0.0010	

It is also of interest to determine the zero field splitting of the ground state. From the spin-Hamiltonian it can be shown that the energies of the three doublets are given by

$$E_{\pm 5/2} = \frac{10}{3}D - \frac{1}{3}a,$$

$$E_{\pm 3/2} = -\frac{2}{3}D + a,$$

$$E_{\pm 1/2} = -\frac{8}{3}D - \frac{2}{3}a.$$

Table III gives the zero field energies at the two temperatures. It is interesting to note that the order of the  $\pm 1/2$  and  $\pm 3/2$  levels reverses in going from room temperature to 90° K.

TABLE III  
Zero field splitting of the ground state of the  
 $Mn^{2+}$  ion in brucite

	295° K		90° K	
	Gauss	$10^{-4} \text{ cm}^{-1}$	Gauss	$10^{-4} \text{ cm}^{-1}$
$E_{\pm 3/2}$	-27.61	-25.91	-49.92	-46.66
$E_{\pm 5/2}$	+15.62	+14.66	+21.93	+20.53
$E_{\pm 1/2}$	+12.00	+11.29	+28.00	+26.13

### DISCUSSION

Certain obvious conclusions can be drawn immediately from the experimental results. The good agreement between experiment and theory confirms the validity of the spin-Hamiltonian for a crystalline field case which results in  $|D| = |a|$ . The absorption spectrum is consistent with trigonal symmetry, which shows that the manganese ions substitute for the magnesium ions in the structure without introducing any significant distortion.

Although it was difficult to make precise measurements of line intensities because of poor resolution, it was found that the intensity ratios agreed with the theoretically predicted ones. Absorption line half-widths were of the order of 10 to 15 gauss. Much narrower line widths have been observed for  $Mn^{2+}$ ; it is most likely that the large widths in the present case are a result of crystal imperfections.

The hyperfine splitting parameters are isotropic and quite typical of results for  $Mn^{2+}$ . On the basis of Van Wieringen's (1955) correlation between covalency and the magnitude of the hyperfine parameters, it can be concluded that brucite is about 90% ionic.

The magnitudes of the crystal field parameters  $D$  and  $a$  can be qualitatively understood on the basis of Pryce's (1950) or Watanabe's (1957) theories. The fact that even for  $D$  small, the two effects, axial and cubic, are distinct and separable, confirms that they must arise through separate mechanisms. The large change in  $D$  with temperature can be interpreted in terms of the large anisotropy in the coefficient of expansion of brucite. The ratio of the coefficient of expansion along the  $c$ -axis to that along the  $a$ -axis is approximately 4:1 (Megaw 1933). One should thus expect large increases in the magnitudes of the axial field and axial field parameter with decreasing temperature.

The  $g$ -value is isotropic but differs by more than the experimental error from the free electron value. The results are, however, typical of  $g$ -values for the  $Mn^{2+}$  ion. There seems to be a slight positive shift in the  $g$ -value at 90° K; this may not be significant because actually the room temperature and low temperature values agree within experimental error.

### ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of this research project by the Ontario Research Foundation. We are also indebted to Dr. R. K. Wanless of the Department of Mines and Technical Surveys, Ottawa, who

kindly supplied the brucite sample that was studied, and to Professor D. M. Shaw of our Geology Department, who performed the spectrographic analysis. W. A. Pieczonka is grateful to the Shell Oil Company of Canada and the Consolidated Mining and Smelting Company of Canada for the financial assistance he received.

## REFERENCES

ABRAGAM, A. and PRYCE, M. H. L. 1951. Proc. Roy. Soc. A, **205**, 135.  
AMINOFF, G. 1919. Geol. Fören. i Stockholm Förh. **41**, 407.  
BETHE, H. A. 1929. Ann. Physik, **3**, 133.  
BLEANEY, B. 1951. Phil. Mag. **42**, 441.  
BLEANEY, B. and INGRAM, D. J. E. 1951. Proc. Roy. Soc. A, **205**, 1369.  
ELLEMAN, D. D. and WILLIAMS, D. 1956. J. Chem. Phys. **25**, 742.  
HIRSHON, J. M. and FRAENKEL, G. K. 1955. Rev. Sci. Instr. **26**, 34.  
LOW, W. 1957. Phys. Rev. **105**, 793.  
MEGAW, H. D. 1933. Proc. Roy. Soc. A, **142**, 198.  
PRYCE, M. H. L. 1950a. Proc. Phys. Soc. A, **63**, 25.  
——— 1950b. Phys. Rev. **80**, 1107.  
RABI, I. I. 1927. Phys. Rev. **29**, 174.  
VAN VLECK, J. H. and PENNEY, W. G. 1934. Phil. Mag. **17**, 961.  
VAN WIERINGEN, J. S. 1955. Discussions Faraday Soc. **19**, 118.  
WATANABE, H. 1957. Progr. Theoret. Phys. **18**, 405.

# MAXIMUM OF THE FACTOR OF ENCIRCLED ENERGY<sup>1</sup>

GUY LANSRAUX AND GERMAIN BOIVIN<sup>2</sup>

## ABSTRACT

It is recognized that the most favorable distribution of radiant energy in a diffraction pattern is that which corresponds to the best concentration around the center  $O$ . This hypothesis is expressed by an extremal condition on the factor of encircled energy  $E(W)$ , that is, the ratio of the energy inside a circle of radius  $W$  and centered on the diffraction pattern, to the total energy in the same.

A study of the effects of spherical aberration on this factor of encircled energy has shown that aberration always tends to decrease the factor from the value obtained with an Airy pattern. However, this factor may be increased by the use of an amplitude filter at the pupil of the optical system.

In treating the case of amplitude filters one may use a rigorous analysis in terms of Taylor's series in  $(1-x^2)^{p-1}$  or a polynomial  $T_n(x)$  of degree  $n-1$  in terms of  $(1-x^2)$ .

The corresponding amplitudes in the diffraction pattern are  $\Gamma(W)$  and  $\Gamma_n(W)$ ; the maximum factor of encircled energy  $E(W_m)$  and  $E_n(W_m)$ . The following convergences are established:  $E_n(W_m) \rightarrow E(W_m)$ ,  $T_n(x) \rightarrow T(x)$ , and  $\Gamma_n(W) \rightarrow \Gamma(W)$  as  $n \rightarrow \infty$ .

When the interval  $(0, W_m)$  of the diffraction pattern is made to correspond to the interval  $(0, 1)$  of the pupil by means of a suitable normalization the amplitude distributions  $T(x)$  and  $\Gamma(W)$ —with  $W = W_m x$ —are identical. Some properties are deduced from this relation; for example, the Airy pattern is the limit of  $\Gamma(W)$  when  $W_m \rightarrow 0$ ; on the other hand, the Gauss function  $e^{-(W_m/2)x^2}$  is an asymptotic expression of  $T(x)$  when  $W_m \rightarrow \infty$ . In any case, the factor of encircled energy is connected to the marginal amplitude in the pupil by the relation  $E(W_m) = 1 - T^2(1)$ .

The numerical determination of  $E(W)$  given up to  $W = 10$  and  $W_m = 2, 3, 4$ , and 5 can be extended by use of an asymptotic expression of the factor of encircled energy.

Finally, a curve  $M(W)$  has been obtained, which is an envelope of the curves  $E(W)$  corresponding to various values of  $W_m$ . This gives the locus of the maximum factor of encircled energy and represents the limiting performance of optical systems.

## INTRODUCTION

The image of a point is never a point but is spread into a diffraction pattern in the observation plane. The consequences of this are well known: limitation of resolution, distortion of the structure of fine details in the image of an extended object, degradation of contrasts. . . . These degrading effects are the more important the larger the diffraction pattern into which the radiant energy is spread. Conversely, the quality of optical images increases as the radiant energy is better concentrated into the central region of the diffraction pattern. Thus, one is allowed to suppose that the most favorable energy distribution is that which corresponds to the greatest concentration of energy in the near vicinity of the center of the diffraction pattern.

As a measure of this concentration of energy we may take the ratio  $E(W)$  of the energy distributed inside a circle of radius  $W$  centered on the focus of the diffraction pattern to the total energy in the pattern (Fig. 1).

<sup>1</sup>Manuscript received May 31, 1960.

Contribution from the Physics Department of Laval University, Quebec, Que.

<sup>2</sup>Holder of a National Research Council Studentship.

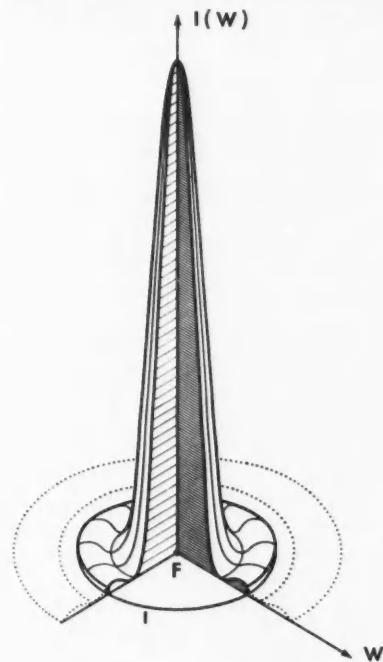


FIG. 1. View in perspective of a diffraction pattern of revolution. The factor of encircled energy is the ratio of the energy inside the circle  $\Gamma$  to the total energy in the diffraction pattern.

To make  $E(W)$  a maximum for a certain radius  $W$  is to gather as much energy as possible into the circle  $\Gamma$  and thereby to optimize the diffraction pattern.

This idea was utilized and studied in brief by Straubel (1935). We have deemed it necessary to enlarge upon and to extend this work.

This function  $E(W)$  has been named by one of us (Lansraux 1946) as the factor of encircled energy by reason of its definition.

For an axially symmetric diffraction pattern of which the radial intensity distribution is  $I(W)$  the factor of encircled energy is

$$(1) \quad E(W) = \frac{\int_0^W I(W)d(W^2)}{\int_0^\infty I(W)d(W^2)}.$$

Whatever the intensity distribution  $I(W)$ , the factor of encircled energy  $E(W)$  vanishes at the origin ( $W = 0$ ), and increases monotonically, approaching unity asymptotically as  $W$  tends to infinity (Figs. 5 to 9).

### I. EFFECTS OF SPHERICAL ABERRATION AND DEFECT OF FOCUS

The absolute maximum of the factor of encircled energy at chosen abscissa  $W = W_m$  is attained only with an optical system which is free of spherical aberration and of defect of focus.

Before proving this statement we shall recall the classical expression for diffraction in optical instruments.

The spherical wave  $\Sigma$  (Fig. 2) emitted by an axial point object  $H$  is transformed into an emergent wavefront  $\Omega$  whose departure from the reference sphere  $S$  may be described by the phase

$$\phi(P) = \frac{2\pi}{\lambda} \overline{PM}.$$

The defect of focus is the distance  $FC$  between the center of curvature  $C$  of the optical system and the intersection  $F$  of the observation plane  $\Pi$  with

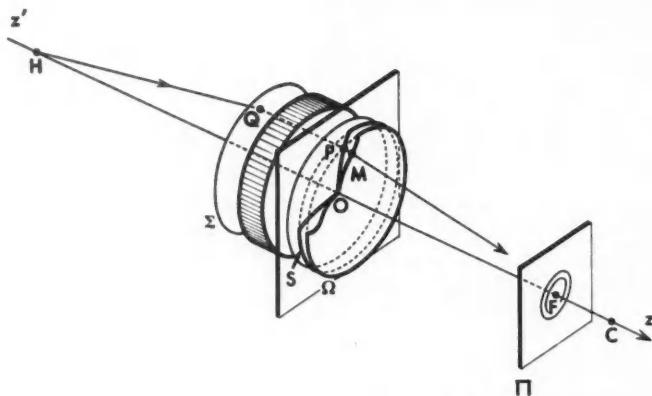


FIG. 2. Schematic diagram of the formation of a diffraction pattern in an optical instrument.

the axis  $Z'Z$  of the optical system. The coefficient of defect of focus is  $\alpha = \pi(\overline{FC}/\overline{OC})$  (Lansraux 1952).

If the amplitude varies from point to point along the wavefront  $\Omega$  by reason of a non-uniform transparency of the optical system for various light paths  $HQPMC$  the transmission coefficient is

$$T(P) = \frac{\text{amplitude at } P}{\text{amplitude at } Q}.$$

For radial co-ordinates  $x$  and  $W$  in the pupil and in the observation plane respectively, the complex amplitude in the diffraction pattern is

$$(2) \quad \gamma(W) = \gamma_r(W) + i\gamma_i(W) = \int_0^1 T(x) e^{i\Psi(x)} J_0(Wx) d(x^2)$$

with  $\Psi(x) = \phi(x) + \alpha x^2$ , and where  $J_0(Wx)$  is the Bessel function of the first kind and of order zero.

We shall now proceed to show that the factor of encircled energy

$$(3) \quad E(W_m) = \frac{\int_0^{W_m} [\gamma_r^2(W) + \gamma_i^2(W)] d(W^2)}{\int_0^{\infty} [\gamma_r^2(W) + \gamma_i^2(W)] d(W^2)}$$

for a system affected by spherical aberration and defect of focus cannot be an absolute maximum.

Consider first two optical systems free of aberration and correctly focused and having transmission functions  $T_r(x) = T(x) \cos \Psi(x)$  and  $T_i(x) = T(x) \sin \Psi(x)$ . Let the amplitude of the respective diffraction patterns be  $\gamma_r(W)$  and  $\gamma_i(W)$  so that

$$(4) \quad E_r(W_m) = \frac{\int_0^{W_m} \gamma_r^2(W) d(W^2)}{\int_0^{\infty} \gamma_r^2(W) d(W^2)} \quad E_i(W_m) = \frac{\int_0^{W_m} \gamma_i^2(W) d(W^2)}{\int_0^{\infty} \gamma_i^2(W) d(W^2)}.$$

All integrals which appear in either numerator or denominator of  $E_r(W_m)$  or of  $E_i(W_m)$  are positive.

Comparing with (3) we have

$$(5) \quad \begin{aligned} E(W_m) &< E_r(W_m) & \text{if} & \quad E_i(W_m) < E_r(W_m), \\ E(W_m) &< E_i(W_m) & \text{if} & \quad E_r(W_m) < E_i(W_m). \end{aligned}$$

Thus for any system with transmission  $T(x)$  and corresponding to either situation it is possible to find a system free of aberration which has a factor of encircled energy greater than that of the aberrant system.

For sake of completeness we must consider the particular case of  $E_r(W_m) = E_i(W_m) = E(W_m)$ . This means that there are three different systems which give the absolute maximum of encircled energy, two of these being free of aberration. But it is shown in Section VI that there is a unique non-aberrant system which yields the absolute maximum of encircled energy, two systems being, however, equivalent, when the amplitudes  $\gamma_i(W)$  and  $\gamma_r(W)$  are proportional. Whence, we may write

$$(6) \quad \gamma_r(W) = C_r g(W), \quad \gamma_i(W) = C_i g(W), \quad \gamma(W) = [C_r + iC_i]g(W).$$

The amplitude  $\gamma(W) = \gamma_r + i\gamma_i$  may thus be written as the product of a real function  $g(W)$  and a complex coefficient  $C_r + iC_i$ . This means that a constant phase change is introduced at the pupil of a system free of aberration and correctly focused. The three systems differ only in this respect.

In consequence of this manifest superiority of aberration-free systems we shall consider in the following parts of this paper only systems which are free of aberration and defect of focus. We shall consider only variation of  $E(W_m)$  which arises from changes in the transmission  $T(x)$  obtained by an amplitude filter (Lansraux 1946, 1953).

II. CONDITION  $T(x) > 0$ 

In principle the function  $T(x)$  should be positive in order that it may represent a transmission coefficient. It will be shown that the extremal condition on  $E(W_m)$  excludes negative values of  $T(x)$  at least for  $W_m < Z = 3.83 \dots$ ,  $Z$  being the first zero of the Airy pattern.

In a paper on the numerical evaluation of  $E(W)$  (Lansraux and Boivin 1958), we have shown that

$$(7) \quad E(W) = \frac{\int_0^1 \int_0^1 T(s)T(t)D(W, s, t)dsdt}{2 \int_0^1 T^2(x)d(x^2)}$$

where  $D(W, s, t)$  is real and positive for  $0 < W < 3.83, \dots$  for all values of  $s$  and  $t$  in the domain of integration  $(0, 1)$ .

If  $T(x)$  changes sign as  $x$  varies from 0 to 1 we may separate this function into positive and negative parts  $T^+$  and  $T^-$  respectively. In abbreviated notation we then have from (7)

$$(8) \quad E(W) = \frac{\iint T^+T^+D}{2\int T^2} + \frac{\iint T^-T^-D}{2\int T^2} + \frac{\iint [T^+T^- + T^-T^+]D}{2\int T^2}.$$

For comparison consider the positive transmission  $t(x) = T^+ - T^-$  with the same total energy transmission. The factor of encircled energy is

$$(9) \quad e(W) = \frac{\iint T^+T^+D}{2\int T^2} + \frac{\iint T^-T^-D}{2\int T^2} - \frac{\iint [T^+T^- + T^-T^+]D}{2\int T^2}.$$

Since  $T^+$  and  $D$  are uniformly positive and  $T^-$  negative it follows that

$$(10) \quad \frac{\iint T^+T^+D}{2\int T^2} > 0, \quad \frac{\iint T^-T^-D}{2\int T^2} > 0, \quad \frac{\iint [T^+T^- + T^-T^+]D}{2\int T^2} < 0.$$

Hence

$$(11) \quad e(W) > E(W).$$

Thus, for any transmission function  $T(x)$  containing negative values the factor of encircled energy can be increased by changing the sign of the negative parts to positive; from this we conclude that the function  $T(x)$  corresponding to the extremal condition must be positive.

This property is proved only for  $W_m < 3.83 \dots$  but the numerical values given in this work for  $W_m = 4$  and  $W_m = 5$  show that it is not restricted to the range  $0 < W < 3.83$ .

It may be noted that the values obtained for  $T(x)$  are all positive; nevertheless, negative values of  $T(x)$  have a physical meaning which shall be obvious if we consider that  $-T(x) = T(x)e^{i\pi}$ . In other words, negative values of  $T(x)$  imply a transmission together with a phase change of  $\pi$ , and may be realized in practice by covering the relevant region of the aperture with a half-wave plate.

### III. ANALYTICAL EXPANSION OF THE PUPIL FUNCTION $T(x)$ AND OF THE DIFFRACTION AMPLITUDE $\gamma(W)$

In his paper of 1935, Straubel made use of a linear expansion in terms of polynomials  $(1-x^2)^{\frac{1}{2}}$ .

One of us used this type of expansion in preceding work (Lansraux 1947). An ambiguous remark of Osterberg and Wilkins (1944) induced Wolf (1951) to question the validity of such expansion. Later Dossier (1954) expressed the doubts about the validity of the expansion. However, one of us (Lansraux 1953) proved conclusively that the representation as given below is quite valid in general.

$$(12) \quad T(x) = \frac{1}{\nu} \sum_{p=1}^{p=\infty} p k_p (1-x^2)^{p-1},$$

$$(13) \quad \gamma(W) = \frac{1}{\nu} \sum_{p=1}^{\infty} k_p L_p(W),$$

where  $\nu$  is a normalization coefficient,  $k_p$  are expansion parameters, and  $L_p$  is a function defined in terms of the Bessel function of the first kind by the relation

$$(14) \quad L_p(W) = p \int_0^1 (1-x^2)^{p-1} J_0(Wx) d(x^2) = 2^p p! \frac{J_p(W)}{W^p}.$$

Further, one of us (Lansraux 1947) has shown that finite expansions may be used with accuracy increasing with order  $n$ .

$$(15) \quad T_n(x) = \frac{1}{\nu} \sum_{p=1}^{p=n} p k_p^n (1-x^2)^{p-1},$$

$$(16) \quad \gamma_n(W) = \frac{1}{\nu} \sum_{p=1}^{p=n} k_p^n L_p(W).$$

In the following paragraphs we obtain the maximum values of the factor of encircled energy  $E_n(W_m)$  and the corresponding distribution  $\gamma_n(W)$  and  $T_n(x)$  in successive order of approximation  $n = 1, n = 2, \dots$ . It is shown that the functions  $E_n(W_m)$ ,  $\gamma_n(W)$ , and  $T_n(x)$  converge respectively to the greatest value  $E(W_m)$  and its corresponding distributions  $\gamma(W)$  and  $T(x)$ .

### IV. MAXIMUM FACTOR OF ENCIRCLED ENERGY $E_n(W_m)$

The factor of encircled energy  $E_n(W)$  calculated from the  $n$ -term expansion (16) of  $\gamma_n(W)$  is

$$(17) \quad E_n(W) = \frac{\int_0^W \gamma_n^2(W) d(W^2)}{\int_0^{\infty} \gamma_n^2(W) d(W^2)} = \frac{\sum_{p=1}^{p=n} \sum_{q=1}^{q=n} k_p^n k_q^n a_p^q(W)}{\sum_{p=1}^{p=n} \sum_{q=1}^{q=n} k_p^n k_q^n a_p^q(\infty)}$$

with

$$a_p^q(W) = \int_0^W L_p(W) L_q(W) d(W^2).$$

The extremal condition on  $E_n(W)$  requires the vanishing of the  $n$  partial derivatives  $\partial E_n(W_m)/\partial k_p^n$  at the chosen abscissa  $W = W_m$ . This gives after calculation

$$(18) \quad \frac{\partial E_n(W_m)}{\partial k_p^n} = 2 \frac{\sum_{q=1}^{q=n} k_p^n [a_p^q(W_m) - E_n(W_m) a_p^q(\infty)]}{\sum_{p=1}^{p=n} \sum_{q=1}^{q=n} k_p^n k_q^n a_p^q(\infty)} = 0.$$

The denominator of the partial derivatives (18) is always finite and non-vanishing since it represents the total energy in the diffraction pattern. Consequently, the maximum value of  $E_n(W_m)$  is the solution of the set of homogeneous equations

$$(19) \quad k_1^n [a_1^1(W_m) - E_n(W_m) a_1^1(\infty)] + \dots + k_n^n [a_n^n(W_m) - E_n(W_m) a_n^n(\infty)] = 0,$$

...

$$k_1^n [a_1^1(W_m) - E_n(W_m) a_1^1(\infty)] + \dots + k_n^n [a_n^n(W_m) - E_n(W_m) a_n^n(\infty)] = 0.$$

No solution of this set—apart from the trivial solution  $k_p^n = 0$ —will exist unless its determinant vanishes

$$(20) \quad \Delta = \begin{vmatrix} a_1^1(W_m) - E_n(W_m) a_1^1(\infty) & \dots & a_1^n(W_m) - E_n(W_m) a_1^n(\infty) \\ \dots & & \dots \\ a_n^1(W_m) - E_n(W_m) a_n^1(\infty) & \dots & a_n^n(W_m) - E_n(W_m) a_n^n(\infty) \end{vmatrix} = 0.$$

This condition is an equation of degree  $n$  in terms of  $E_n(W_m)$  and its  $n$  roots are extrema of the factor of encircled energy. We are here interested in the greatest of these roots. Substitution of this value of  $E_n(W_m)$  into equations (19) yields a set of proportional values of  $k_1^n \dots k_n^n$  one of these parameters being left undetermined. In order that the solution be completely determined it is necessary to impose a supplementary normalizing condition which, of course, does not limit the generality of the solution. In order to compare pupil function  $T_n(x)$  for successive orders  $n$  for maxima of  $E_n(W_m)$  at various abscissas  $W_m$  it is convenient to choose the normalization so that the pupil function is unity at the origin:  $T_n(0) = 1$ .

The corresponding amplitudes in the diffraction pattern  $\gamma_n(W)$  are not then normalized since they are related to  $T_n(x)$  through a Hankel integral analogous to relation (2). Therefore in order to compare diffraction patterns it is further convenient to introduce normalized amplitudes  $\Gamma_n(W) = \nu \gamma_n(W)$  with normalizing coefficient  $\nu$  fixed by the condition of unit amplitude at the origin  $\Gamma_n(0) = 1$ .

In order to avoid any possible confusion we list our conventions as follows:  
Normalized amplitude in the pupil

$$(21) \quad T_n(x) = \frac{1}{\nu} \sum_{p=1}^{p=n} p k_p (1-x^2)^{p-1}$$

with

$$T_n(0) = \frac{1}{\nu} \sum_{p=1}^{p=n} p k_p = 1.$$

Amplitude in the diffraction pattern corresponding to the normalized pupil function

$$(22) \quad \begin{aligned} \gamma_n(W) &= \int_0^1 T_n(x) J_0(Wx) d(x^2), \\ \gamma_n(W) &= \frac{1}{\nu} \sum_{p=1}^{p=n} k_p L_p(W). \end{aligned}$$

Normalized amplitude in the diffraction pattern

$$(23) \quad \begin{aligned} \Gamma_n(W) &= \nu \gamma_n(W) = \sum_{p=1}^{p=n} k_p L_p(W) \\ \text{with} \quad \Gamma_n(0) &= \nu \gamma_n(0) = \sum_{p=1}^{p=n} k_p = 1. \end{aligned}$$

Normalization coefficient

$$(24) \quad \nu = \frac{\Gamma_n(W)}{\gamma_n(W)} = \frac{1}{\gamma_n(0)} = \sum_{p=1}^{p=n} p k_p.$$

It is important to note at this point that the greatest value of  $E_n(W_m)$  uniquely determines the normalized amplitudes  $T_n(x)$ ,  $\Gamma_n(W)$ , and their expansion coefficients  $k_p^n$ . In particular, we shall use the following property: any function of order  $n$ , satisfying the integral equation

$$(25) \quad \int_0^{W_m} \delta_n^2(W) d(W^2) - E_n(W_m) \int_0^\infty \delta_n^2(W) d(W^2) = 0$$

must be proportional to  $\Gamma_n(W)$  for any  $W$ .

## V. CONVERGENCE OF EXTREMAL VALUES OF $E_n(W_m)$ FOR INCREASING $n$

The extremal values of  $E_n(W_m)$  converge to a limit  $E(W_m)$  when  $n \rightarrow \infty$  because the set of extremal values  $E_1(W_m), \dots, E_{n-1}(W_m), E_n(W_m)$  constitutes a sequence of increasing positive and finite terms.

The sequence of extremal values  $E_n(W_m)$  is certainly increasing since the inequality  $E_{n-1}(W_m) > E_n(W_m)$  is absurd. Any  $n-1$  term expansion is a special case of a  $n$ -term expansion with  $k_n^n = 0$  and cannot give a value of the factor of encircled energy greater than the extremum of  $E_n(W_m)$ .

Also the uniquely determined extremal values of  $E_n(W_m)$  imply a uniquely determined sequence  $E_1(W_m), E_2(W_m), \dots, E_n(W_m)$  and a unique limit  $E(W_m)$ .

Table I contains a set of values of  $E_n(W_m)$  and differences  $\Delta E_n(W_m) = E_n(W_m) - E_{n-1}(W_m)$ . It exhibits the rapid convergence of the extremals of the factor of encircled energy with increasing order  $n$ . It may be seen that the convergence is more rapid for smaller values of  $W_m$ .

TABLE I

Absolute maximum  $E_n(W_m)$  of the factor of encircled energy and the differences  $\Delta E_n(W_m)$  with respect to the order  $n$

$n$	$E_n(2)$	$\Delta E_n(2)$	$E_n(3)$	$\Delta E_n(3)$	$E_n(4)$	$\Delta E_n(4)$	$E_n(5)$	$\Delta E_n(5)$
1	0.617261	0.012347	0.817412	0.068977	0.837910	0.132463	0.861151	0.121334
2	0.629608	0.000022	0.886389	0.000660	0.970373	0.004547	0.982485	0.012630
3	0.629630		0.887049	0.000001	0.974920	0.000031	0.995115	0.000226
4			0.887050		0.974951		0.995341	
5							0.995342	0.000001

### VI. CONVERGENCE OF THE FUNCTIONS $\Gamma_n(W)$ AND $T_n(x)$ WITH INCREASING $n$

The convergence of  $E_n(W_m)$  implies that of  $\Gamma_n(W)$  and of  $T_n(x)$ .

Consider two normalized functions  $\Gamma_{n-1}(W)$ ,  $\Gamma_n(W)$  of consecutive orders  $n-1$ ,  $n$  and having respective total energies

$$(26) \quad \tau_{n-1} = \int_0^\infty \Gamma_{n-1}^2(W) d(W^2) \quad \tau_n = \int_0^\infty \Gamma_n^2(W) d(W^2).$$

The functions

$$\frac{\Gamma_{n-1}(W)}{\sqrt{\tau_{n-1}}}, \quad \frac{\Gamma_n(W)}{\sqrt{\tau_n}}, \quad \text{and} \quad \delta_n(W) = \frac{\Gamma_n(W)}{\sqrt{\tau_n}} - \frac{\Gamma_{n-1}(W)}{\sqrt{\tau_{n-1}}}$$

have the following properties:

$$(27) \quad \int_0^W \left[ \frac{\Gamma_{n-1}(W)}{\sqrt{\tau_{n-1}}} \right]^2 d(W^2) = E_{n-1}, \quad \int_0^W \left[ \frac{\Gamma_n(W)}{\sqrt{\tau_n}} \right]^2 d(W^2) = E_n(W),$$

$$(28) \quad \int_0^{W_m} \delta^2(W) d(W^2) = E_n(W_m) + E_{n-1}(W_m) - \frac{2}{\sqrt{\tau_n \tau_{n-1}}} \int_0^{W_m} \Gamma_n(W) \Gamma_{n-1}(W) d(W^2),$$

$$(29) \quad \int_0^\infty \delta^2(W) d(W^2) = 2 - \frac{2}{\sqrt{\tau_n \tau_{n-1}}} \int_0^\infty \Gamma_n(W) \Gamma_{n-1}(W) d(W^2).$$

But the set of equations (19) from which  $E_n(W_m)$  and  $\Gamma_n(W)$  are obtained is equivalent to

$$(30) \quad \int_0^{W_m} \Gamma_n(W) L_p(W) d(W^2) - E_n(W_m) \int_0^\infty \Gamma_n(W) L_p(W) d(W^2) = 0$$

with  $p = 1, \dots, n$ .

From a linear combination of relation (30) one obtains

$$(31) \quad \int_0^{W_m} \Gamma_n(W) \Gamma_{n-1}(W) d(W^2) - E_n(W_m) \int_0^\infty \Gamma_n(W) \Gamma_{n-1}(W) d(W^2) = 0,$$

whence the relation between integrals (28) and (29)

$$(32) \quad \int_0^{W_m} \delta_n^2(W) d(W^2) - E_n(W_m) \int_0^\infty \delta_n^2(W) d(W^2) = E_{n-1}(W) - E_n(W).$$

The convergence of  $E_n$  when  $n \rightarrow \infty$  implies

$$(33) \quad \lim_{n \rightarrow \infty} \int_0^{W_m} \delta_n^2(W) d(W^2) - E_n(W_m) \int_0^{\infty} \delta_n^2(W) d(W^2) = 0.$$

Now, from (25) the solution of (23) can only be the product of  $\Gamma_n(W)$  by a constant  $C$ . That is,

$$(34) \quad \lim_{n \rightarrow \infty} \delta_n(W) = \lim_{n \rightarrow \infty} \frac{\Gamma_n(W)}{\sqrt{\tau_n}} - \frac{\Gamma_{n-1}(W)}{\sqrt{\tau_{n-1}}} = \lim_{n \rightarrow \infty} C \Gamma_n(W).$$

The constant  $C$  is determined by the normalization conditions  $\Gamma_{n-1}(0) = \Gamma_n(0) = 1$  and one finds

$$(35) \quad \lim_{n \rightarrow \infty} [\Gamma_n(W) - \Gamma_{n-1}(W)] = 0.$$

This relation establishes the convergence of  $\Gamma_n(W)$  to a limit of  $\Gamma(W)$ .

Expanding (35)

$$(36) \quad \lim_{n \rightarrow \infty} \left[ k_n^n L_n(W) + \sum_{p=1}^{p=n-1} (k_p^n - k_p^{n-1}) L_p(W) \right] = 0.$$

It is concluded that  $k_n^n \rightarrow 0$  and  $k_p^n - k_p^{n-1} \rightarrow 0$  when  $n \rightarrow \infty$ . Consequently, all coefficients  $k_p^n \rightarrow 0$  when  $p$  and  $n$  independently increase and approach infinity. This is a necessary condition for the convergence of limit function

$$\Gamma(W) = \sum_{p=1}^{p=\infty} k_p L_p(W).$$

All these properties are apparent in the numerical values of  $k_p^n$  shown in Table II.

TABLE II  
Coefficients  $k_p^n$  in the  $n$ -term expansion of normalized amplitudes  
 $\Gamma_n(W)$  corresponding to the absolute maximum  $E_n(W_m)$

$$\Gamma_n(W) = \sum_{p=1}^{p=n} k_p^n L_p(W) \quad \text{with} \quad \Gamma_n(0) = \sum k_p^n = 1$$

$n$	$k_1^n$	$k_2^n$	$k_3^n$	$k_4^n$	$k_5^n$
$W_m = 2$					
1	1.00000				
2	0.75370	0.24630			
3	0.76730	0.20550	0.02721		
$W_m = 3$					
1	1.00000				
2	0.47544	0.52456			
3	0.53868	0.33365	0.12767		
4	0.53514	0.35488	0.09229	0.01769	
$W_m = 4$					
1	1.00000				
2	0.17725	0.82275			
3	0.33628	0.31954	0.34418		
4	0.31969	0.41862	0.17912	0.08257	
$W_m = 5$					
1	1.00000				
2	0.00285	0.99715			
3	0.21643	0.11841	0.66516		
4	0.16718	0.40436	0.18825	0.24022	
5	0.17124	0.36367	0.31031	0.09782	0.05696

The convergence of  $\Gamma_n(W)$  implies that of the corresponding pupil functions  $T_n(x)$ . Figures 3 and 4 show the course of the functions  $T_n(x)$  for various values of  $W_m$ . Also, the second diagram of Fig. 4 exhibits the convergence of the plots of the factor of encircled energy  $E_n(W)$  for various order  $n$  and for  $W_m = 5$ .

All those curves are plotted from the numerical values of Table III.

TABLE III

Normalized amplitudes  $T_n(x)$  corresponding to the absolute maximum  $E_n(W_m)$ . Limit values of the amplitude  $T(x)$

$$T_n(x) = \frac{1}{\sum p k_p^n} (1-x^2)^{p-1} \quad \text{with} \quad T_n(0) = 1$$

$x$	$T_1(x)$	$T_2(x)$	$T_3(x)$	$T_4(x)$	$T_5(x)$	$T(x)$
$W_m = 2$						
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
0.2	1.0000	0.9482	0.9819	0.9819	0.982	0.982
0.4	1.0000	0.9368	0.9287	0.9287	0.929	0.929
0.6	1.0000	0.8577	0.8443	0.8443	0.844	0.844
0.8	1.0000	0.7478	0.7348	0.7348	0.735	0.735
1.0	1.0000	0.6048	0.6090	0.6090	0.609	0.609
$W_m = 3$						
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
0.2	1.0000	0.9725	0.9643	0.9634	0.9634	0.963
0.4	1.0000	0.8899	0.8618	0.8594	0.8594	0.859
0.6	1.0000	0.7523	0.7065	0.7041	0.7041	0.704
0.8	1.0000	0.5596	0.5214	0.5211	0.5211	0.521
1.0	1.0000	0.3119	0.3390	0.3360	0.3360	0.336
$W_m = 4$						
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
0.2	1.0000	0.9639	0.9470	0.9438	0.9438	0.944
0.4	1.0000	0.8556	0.7977	0.7892	0.7892	0.789
0.6	1.0000	0.6750	0.5818	0.5740	0.5740	0.574
0.8	1.0000	0.4222	0.3487	0.3488	0.3488	0.349
1.0	1.0000	0.0972	0.1675	0.1579	0.1579	0.158
$W_m = 5$						
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
0.2	1.0000	0.9601	0.9322	0.9251	0.9241	0.924
0.4	1.0000	0.8402	0.7446	0.7254	0.7235	0.724
0.6	1.0000	0.6405	0.4841	0.4669	0.4663	0.466
0.8	1.0000	0.3609	0.2288	0.2304	0.2302	0.230
1.0	1.0000	0.0014	0.0884	0.0668	0.0683	0.068

We note, also, that the uniqueness of values  $E_1(W), \dots, E_n(W), \dots, E(W)$  implies the uniqueness of the sequences  $\Gamma_1(W), \dots, \Gamma_n(W), \dots, \Gamma(W)$  and  $T_1(x), \dots, T_n(x), \dots, T(x)$ .

In particular there exists but one normalized pupil function  $T(x)$  and but one normalized diffraction pattern which yield the absolute factor of encircled energy  $E(W_m)$ .

Expression (6) for the functions  $\gamma_r(W)$  and  $\gamma_i(W)$  satisfying the integral equation (4), Section I, are thus justified.

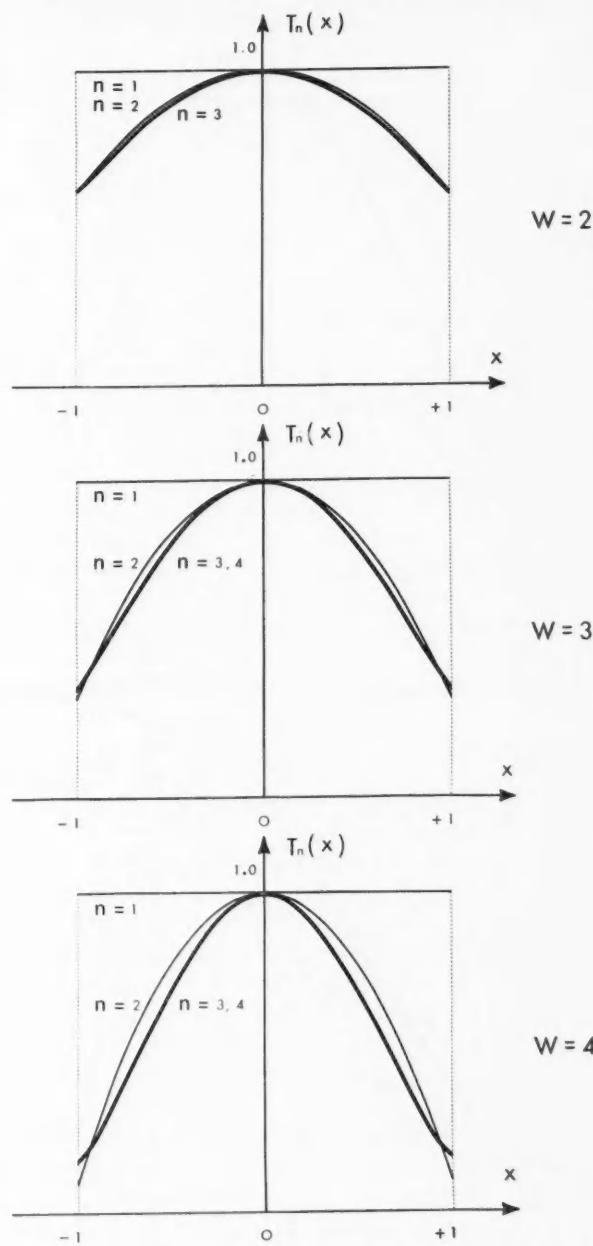


FIG. 3. Curves of normalized amplitudes  $T_n(x)$  for various order  $n$  and  $W_m = 2, 3, 4$ .

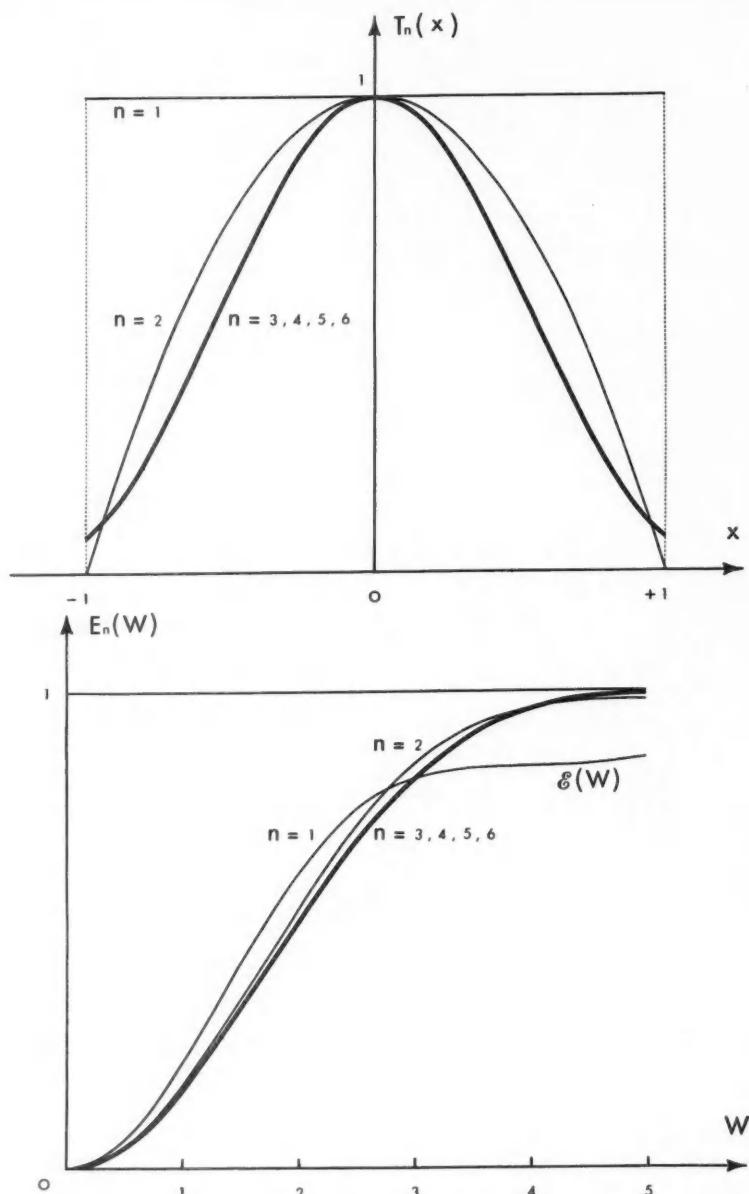


FIG. 4. Curves of normalized amplitudes  $T_n(x)$  and  $E_n(W)$  for  $W_m = 5$  and  $n = 1, 2, 3, 4, 5, 6$ .

VII. AN INTERESTING PROPERTY OF THE LIMIT FUNCTIONS  $T(x)$  AND  $\Gamma(W)$ 

By means of suitable normalization the light distribution in the pupil is identical with that of the diffraction patterns inside the circle of radius  $W_m$  centered at the focus. Analytically that property can be written

$$(37) \quad \begin{aligned} T(x) &= \Gamma(W_m x) & 0 < x < 1 \\ \Gamma(W) &= T\left(\frac{W}{W_m}\right) & 0 < W < W_m \end{aligned} \quad \text{with } x = \frac{W}{W_m}.$$

To prove this, we will assume that the portion of the diffraction pattern inside the circle of radius  $W_m$  is a new pupil whose diffraction pattern is

$$(38) \quad \theta(x) = \int_0^{W_m} \gamma(W) J_0(Wx) d(W^2).$$

It will be proved that  $\theta(x)$  is proportional to  $T(x)$  in the interval  $0 < x < 1$ . Applying Parseval's theorem to the pairs of transforms  $T(x)$  and  $\gamma(W)$ ,  $\theta(x)$  and  $\gamma(W)$ ,  $(1-x^2)^{p-1}$ , and  $L_p(W)/p$  we have

$$(39) \quad \int_0^1 T(x) (1-x^2)^{p-1} d(x^2) = \frac{1}{4p} \int_0^\infty \gamma(W) L_p(W) d(W^2),$$

$$(40) \quad \int_0^1 \theta(x) (1-x^2)^{p-1} d(x^2) = \frac{1}{p} \int_0^{W_m} \gamma(W) L_p(W) d(W^2).$$

The extremal condition  $\partial E(W_m)/\partial k_p = 0$  can be written

$$(41) \quad \int_0^{W_m} \gamma(W) L_p(W) d(W^2) - E(W_m) \int_0^\infty \gamma(W) L_p(W) d(W^2) = 0.$$

Hence

$$(42) \quad \int_0^1 [\theta(x) - 4E(W_m) T(x)] (1-x^2)^{p-1} d(x^2) = 0.$$

That integral can be satisfied whatever  $p$  may be, if

$$(43) \quad \theta(x) = 4E(W_m) T(x).$$

Plancherel's theorem applied to  $\gamma(W)$  and  $\theta(x)$ ,  $\gamma(W)$  and  $T(x)$  gives

$$(44) \quad \begin{aligned} \int_0^{W_m} \gamma^2(W) d(W^2) &= \frac{1}{4} \int_0^\infty \theta^2(x) d(x^2), \\ \int_0^\infty \gamma^2(W) d(W^2) &= 4 \int_0^1 T^2(x) d(x^2). \end{aligned}$$

Using (43) which connects  $T(x)$  and  $\theta(x)$ , we have

$$(45) \quad \int_0^\infty \gamma^2(W) d(W^2) = \frac{1}{4E^2(W_m)} \int_0^1 \theta^2(x) d(x^2).$$

Hence

$$(46) \quad E(W_m) = \frac{\int_0^{W_m} \gamma^2(W) d(W^2)}{\int_0^{\infty} \gamma^2(W) d(W^2)} = E^2(W_m) \frac{\int_0^{\infty} \theta^2(x) d(x^2)}{\int_0^1 \theta^2(x) d(x^2)},$$

$$(47) \quad \frac{\int_0^1 \theta^2(x) d(x^2)}{\int_0^{\infty} \theta^2(x) d(x^2)} = E(W_m) = E_{\theta}(1).$$

Since  $\Gamma(W)$  is unique and the auxiliary pupil  $\theta(x)$  gives a factor of encircled energy equal to the absolute maximum, both  $\theta(x)$  and  $\Gamma(W)$  must be identical (apart from a normalization factor); but  $\theta(x)$  and  $T(x)$  being proportional for  $0 < x < 1$ ,  $T(x) = \Gamma(W_m x)$  as shown in Tables III and V.

We can conclude that the differences  $D_n(x) = T_n(x) - \Gamma_n(W_m x) \rightarrow 0$  when  $n \rightarrow \infty$ . Some calculated values of  $D_n(x)$  are given in Table IV for the case  $W_m = 5$  to show how close the expansion of  $T_n(x)$  and  $\Gamma_n(W)$  are to their limits  $T(x)$  and  $\Gamma(W)$  respectively.

The physical meaning of that property has to be mentioned. The pupil  $T(x)$  ( $x < 1$ ) gives a diffraction pattern  $\gamma(W)$  spread out to infinity in the observation plane II. If we put a circular diaphragm of radius  $W_m$  centered on the optical axis, a new diffraction occurs through that new pupil which is identical with  $T(x)$  for  $W < W_m$  by means of a proper normalization of the abscissas and amplitudes.

More generally, an arbitrary number of diffractions can be produced without changing the light distribution in each successive diffraction pattern.

Besides that, it is possible to get the same geometrical dimension in the pupil and in the diffraction pattern. It is known that a dilatation or a contraction of the pupil by a constant  $k$  is followed by a contraction or a dilatation of the diffraction pattern by the constant  $1/k$ . Consequently, there exists a constant  $k = W_m = R$  so that the intervals  $(0, 1)$  and  $(0, W_m)$  of the pupil and diffraction pattern respectively can be transformed into the intervals  $(0, R)$  and  $(0, R)$ .

That normalization of  $x$  and  $W$  will be better expressed by a slight modification in our convention. Let  $x_n$  be the radial co-ordinate of the diffraction pattern lying on the observation plane  $\Pi_n$ . The light distribution included in the circular domain  $0 < x < R$  will be considered as a pupil of order  $n$ .

According to this new convention two pupils of successive order  $n$  and  $n-1$  have proportional amplitudes  $A_n(x_n)$ ,  $A_{n-1}(x_{n-1})$  when  $x_n = x_{n-1}$ .

The transformation connecting  $\theta(x)$  and  $T(x)$  shows that the constant of proportionality between  $A_n(x_n)$  and  $A_{n-1}(x_{n-1})$  is the square root of the factor of encircled energy

$$(48) \quad A_n(x_n) = \sqrt{E(R)} A_{n-1}(x_{n-1}) \quad \text{with} \quad x_n = x_{n-1}.$$

TABLE IV  
Differences  $D_n(x) = T_n(x) - \Gamma_n(W_m x)$

$x$	$T_2(x)$	$\Gamma_2(5x)$	$D_2(x)$	$T_3(x)$	$\Gamma_3(5x)$	$D_3(x)$	$T_4(x)$	$\Gamma_4(5x)$	$D_4(x)$	$T_5(x)$	$\Gamma_5(5x)$	$D_5(x)$
0.0	1.00000	1.00000	0.00000	1.00000	0.00000	1.00000	1.00000	0.00000	0.00000	1.00000	0.00000	0.00000
0.2	0.96006	0.91912	+0.04094	0.93224	0.92394	+0.00830	0.92509	0.92406	+0.00103	0.92414	0.92407	+0.00007
0.4	0.84023	0.70530	+0.13493	0.74462	0.72299	+0.02163	0.72536	0.72340	+0.00196	0.72349	0.72339	+0.00010
0.6	0.64051	0.43149	+0.20902	0.48406	0.46555	+0.01851	0.46690	0.46630	+0.00060	0.46631	0.46630	+0.00001
0.8	0.36091	0.18145	+0.17946	0.22881	0.22901	-0.00020	0.23040	0.23013	+0.00027	0.23019	0.23014	-0.00005
1.0	0.00143	0.01449	-0.01306	0.08838	0.06659	-0.02179	0.06683	0.06823	-0.00140	0.06834	0.06825	-0.00009

TABLE V  
Limit values of the amplitudes  $\Gamma(W)$  and  $T(x)$  corresponding to the absolute maximum  $E(W_m)$

$W$	$W_m = 0$	$W_m = 2$	$W_m = 3$	$W_m = 4$	$W_m = 5$
0.0	+1.000	+1.000	+1.000	+1.000	+1.000
0.5	+0.969	+0.972	+0.974	+0.978	+0.981
1.0	+0.880	+0.890	+0.901	+0.913	+0.924
1.5	+0.744	+0.764	+0.787	+0.812	+0.836
2.0	+0.577	+0.609	+0.645	+0.686	+0.723
2.5	+0.398	+0.441	+0.490	+0.545	+0.597
3.0	+0.226	+0.277	+0.336	+0.403	+0.466
3.5	+0.079	+0.134	+0.198	+0.271	+0.341
4.0	-0.033	+0.021	+0.084	+0.158	+0.230
4.5	-0.103	-0.055	+0.002	+0.070	+0.138
5.0	-0.131	-0.094	-0.048	+0.009	+0.068
5.5	-0.124	-0.100	-0.068	-0.026	+0.021
6.0	-0.092	-0.081	-0.064	-0.039	-0.007
6.5	-0.047	-0.048	-0.046	-0.036	-0.018
7.0	-0.001	-0.012	-0.020	-0.023	-0.018
7.5	+0.036	+0.020	+0.005	-0.007	-0.012
8.0	+0.059	+0.041	+0.024	+0.007	-0.003
8.5	+0.064	+0.049	+0.033	+0.017	+0.004
9.0	+0.055	+0.044	+0.033	+0.020	+0.009
9.5	+0.034	+0.030	+0.025	+0.018	+0.010
10.0	+0.009	+0.011	+0.012	+0.011	+0.008

That relation can be justified by considering the definition of the factor of encircled energy and the conservation of energy between the pupil of order  $n-1$  and the diffraction pattern of order  $n$  (Lansraux 1953); that is,

$$(49) \quad \text{energy of pupil } n = E(R) \times \text{energy of diffraction pattern } n,$$

$$\text{energy of pupil } n-1 = \text{energy of diffraction pattern } n,$$

$$\frac{\text{energy of pupil } n}{\text{energy of pupil } n-1} = E(R) = \frac{\int_0^R A_n^2(x_n) d(x_n^2)}{\int_0^R A_{n-1}(x_{n-1}) d(x_{n-1}^2)}.$$

Or shortly,

$$(50) \quad A_n(x_n) = \sqrt{E(R)} A_{n-1}(x_{n-1}) = \dots = [E(R)]^{(n-1)/2} A_1 x_1 = [E(R)]^{n/2} T(x)$$

with  $x_n = x_{n-1} = \dots = x_1 = x$ .

That interesting property will be the matter of another paper on the limiting performance of optical instruments.

### VIII. SOME PROPERTIES OF $T(x)$ AND $\Gamma(W)$ DEDUCED FROM THE RELATION $T(x) = \Gamma(W_m x)$

The properties derived from the relation  $T(x) = \Gamma(W_m x)$  will be the matter of the present section.

(a) *The normalization coefficient is*

$$(51) \quad \nu = W_m / 2 \sqrt{E(W_m)}.$$

Equations (23), (44), and  $T(x) = \Gamma(W_m x)$  give successively

$$(52) \quad E(W_m) = \frac{\int_0^{W_m} \gamma^2(W) d(W^2)}{\int_0^\infty \gamma^2(W) d(W^2)} = \frac{\frac{1}{\nu^2} \int_0^{W_m} \Gamma^2(W) d(W^2)}{4 \int_0^1 T^2(x) d(x^2)}$$

$$= \frac{W_m^2 \int_0^1 \Gamma(W_m x) d(x^2)}{4\nu^2 \int_0^1 T^2(x) d(x^2)} = \frac{W_m^2}{4\nu^2},$$

which proves (51). Some numerical values of  $\nu$  and  $\gamma(0) = 1/\nu$  are given in Table VI.

TABLE VI

Coefficient of normalization, amplitude at the center of the diffraction pattern, coefficient of total transmission, and intensity relative to a unit total transmission

$W_m$	0	2	3	4	5
$\nu$	1.000	1.260	1.593	2.025	2.506
$\gamma(0)$	1.000	0.794	0.628	0.494	0.399
$\tau$	1.000	0.643	0.431	0.302	0.229
$\gamma(0)/\tau$	1.000	0.980	0.914	0.808	0.694

It can be seen that  $\nu$  increases with  $W_m$  while  $\gamma(0)$ , the central amplitude of the diffraction pattern decreases. The total transmission  $\tau = \int_0^1 T^2(x) d(x^2)$  decreases as  $W_m$  increases. The ratio  $\gamma^2(0)/\tau$ , which represents the intensity relative to a unit total energy, is always smaller than one. In other words, the central intensity  $\gamma^2(0)$  is smaller than  $\tau \cdot A^2(0) = \tau$ , which corresponds to the amplitude at the center of an Airy pattern produced by a uniformly transparent system having the same total transmission  $\tau$ .

(b) The maximum factor of encircled energy  $E(W_m)$  is related to the margin amplitude  $T(1)$  by

$$(53) \quad E(W_m) = 1 - T^2(1).$$

Using the extremal conditions  $\partial E(W_m)/\partial k_p$  expressed by the equivalent relation

$$(54) \quad \int_0^{W_m} \Gamma(W) L_p(W) d(W^2) - E(W_m) \int_0^\infty \Gamma(W) L_p(W) d(W^2) = 0$$

a linear combination leads to

$$\sum_{p=2}^{p=\infty} \frac{-k_{p-1}}{4p} \left[ \int_0^{W_m} \Gamma(W) L_p(W) d(W^2) - E(W_m) \int_0^\infty \Gamma(W) L_p(W) d(W^2) \right] = 0,$$

$$(55) \quad \int_0^{W_m} \Gamma(W) \frac{d\Gamma(W)}{d(W^2)} d(W^2) - E(W_m) \int_0^\infty \Gamma(W) \frac{d\Gamma(W)}{d(W^2)} d(W^2) = 0,$$

$$\frac{1}{2} [\Gamma^2(W_m) - 1] + \frac{1}{2} E(W_m) = 0.$$

Hence

$$(56) \quad E(W_m) = 1 - \Gamma^2(W_m) = \left[ 1 - \Gamma^2(W_m x) \right]_{x=1} = 1 - T^2(1).$$

This expression shows that the larger the factor of encircled energy, the smaller the margin transmission. This agrees very well with the properties of margin area of the diffraction patterns where the intensity is asymptotically proportional to the margin amplitude of the pupil (Lansraux 1953).

(c) From the fact that  $T(x) > 0$  for  $0 < x < 1$  it can be shown that the amplitude is maximum at  $x = 0$  and  $W = 0$  in the pupil and diffraction pattern respectively. It is known that  $J_0(Wx) < 1$  so that

$$(57) \quad |\gamma(W)| = \left| \int_0^1 T(x) J_0(Wx) d(x^2) \right| < \int_0^1 T(x) d(x^2) = \gamma(0),$$

and

$$(58) \quad |\gamma(W)| < \gamma(0) \rightarrow |\Gamma(W)| < \Gamma(0) \rightarrow |\Gamma(W_m x)| < \Gamma(0) \rightarrow T(x) < T(0).$$

This property explains why the normalizations  $T(0) = \Gamma(0) = 1$  were chosen.

(d) The amplitude in the pupil decreases as  $x$  varies from 0 to 1, at least for  $W_m < 3.83 \dots$

The derivative of  $\gamma(W)$  is

$$(59) \quad \frac{d\gamma(W)}{d(W)} = \frac{d}{d(W)} \int_0^1 T(x) J_0(Wx) d(x^2) = -\frac{W}{2} \int_0^1 x T(x) L_1(Wx) d(x^2)$$

which is negative for  $W < W_m = 3.83 \dots$  The same holds for

$$\frac{dT}{dx} = \frac{d\Gamma(W_m x)}{dx} = W_m \frac{d\Gamma(W)}{dW} = \frac{W_m}{\nu} \frac{d\gamma(W)}{d(W)}.$$

The negative values of the derivatives of  $T(x)$  and  $\Gamma(W)$  mean a decreasing amplitude in the pupil and diffraction pattern with increasing radial coordinates.

(e) The amplitude is stationary at the origin of both light distributions. The property is evident from (d) since

$$\frac{dT(0)}{dx} = \frac{d\gamma(0)}{dW} = \frac{d\Gamma(0)}{dW} = 0.$$

(f) When  $W_m \rightarrow 0$  the diffraction pattern  $\Gamma(W)$  corresponding to the absolute factor of encircled energy becomes an Airy pattern.

From (e), the amplitude  $\Gamma(W)$  is stationary for  $0 < W < W_m$  when  $W_m \rightarrow 0$ ; the corresponding  $T(x) = \Gamma(W_m x)$  becomes

$$\lim_{W_m \rightarrow 0} T(x) = \lim_{W_m \rightarrow 0} \Gamma(W_m x) = \Gamma(0) = 1,$$

which is typical of the uniform transparency system producing the Airy pattern.

Even if the Airy pattern is well known we give as a reference the numerical values of its amplitude  $A(W)$  and its factor of encircled energy  $\mathcal{E}(W)$  in Tables V and VIII.

In this case the normalization coefficient  $\nu = 1$ .

(g) For small values of  $W_m$  one can use approximate expressions

$$(60) \quad T(x) \sim 1 - \frac{W_m^2}{8} x^2 \quad E(W_m) \sim \frac{W_m^2}{4} \quad \nu \sim 1.$$

The stationary character of amplitude distributions in the pupil and diffraction pattern (paragraph f) for small values of  $W_m$  justifies the approximate expressions of  $T(x)$ ,  $\gamma(W)$ , and  $\Gamma(W)$ , by means of the two-term expansions (21), (22), (23). From these one can easily deduce the equations mentioned above.

(h) When  $W_m$  increases the amplitude  $T(x)$  corresponding to the absolute maximum of the factor of encircled energy becomes a delta function. In fact, the decrease of  $\Gamma(W)$  for large  $W$  allows to write  $\Gamma(W_m x) \rightarrow 0$  as  $W_m \rightarrow \infty$  whenever  $x \neq 0$ , whence  $T(x) = \Gamma(W_m x) \rightarrow 0$ . However, at the center of the pupil, the normalization  $T(0) = 1$  still holds. Thus  $\lim_{W_m \rightarrow \infty} T(x) = \text{delta function}$ .

(i) For large  $W_m$  the amplitude distribution  $T(x)$  may be approximately represented by the Laplace-Gauss exponential  $T(x) \simeq e^{-(W_m/2)x^2}$ .

This approximation is derived from the fact that the exponential  $e^{-kx^2}$  satisfies the condition  $T(x) = \Gamma(W_m x)$  as a special case, that is

$$(61) \quad \int_0^\infty e^{-kx^2} J_0(Wx) d(x^2) = \frac{1}{k} e^{-W^2/4k} \quad \text{with } k > 0.$$

But for large  $k$  the integrand is small for  $x > 1$ , and the expression becomes

$$(62) \quad \int_0^1 e^{-kx^2} J_0(Wx) d(x^2) \simeq \frac{1}{k} e^{-W^2/4k}.$$

Hence

$$(63) \quad T(x) \simeq e^{-kx^2}, \quad \gamma(W) \simeq \frac{1}{k} e^{-W^2/4k}, \quad \Gamma(W) = e^{-W^2/4k}, \quad \nu \simeq k.$$

The constant  $k = W_m/2$  is determined from the condition  $T(x) = \Gamma(W_m x)$ . Thus we finally have

$$(64) \quad T(x) = \Gamma(W_m x) \simeq e^{-(W_m/2)x^2} \quad \text{and} \quad \nu \simeq \frac{W_m}{2}.$$

By comparison with the expression  $\nu = W_m/2\sqrt{E(W_m)}$  the approximation implies that  $E(W_m) \simeq 1$  which get better the larger  $W_m$ . For example in the case  $W_m = 5$  for which  $E(W_m) = 0.9953$  the departure of  $e^{-2.5x^2}$  from  $T(x)$  is better than 0.06 as Table VII shows; moreover, the larger  $W$ , the better the approximation.

TABLE VII

Compared values of the limit function  $T(x)$  and of the exponential  $e^{-2.5x^2}$  for  $W_m = 5$

$x$	0.0	0.2	0.4	0.6	0.8	1.0
$T(x)$	1.000	0.924	0.723	0.466	0.230	0.068
$e^{-2.5x^2}$	1.000	0.905	0.670	0.407	0.202	0.082

Finally the delta function can be derived again from  $T(x)$  (equation 64), which confirms paragraph (h).

### IX. ASYMPTOTIC EXPRESSION $e(W)$ OF THE FACTOR OF ENCIRCLED ENERGY $E(W)$

An approaching expression of  $\gamma(W)$  can be obtained for large  $W$  when the expansion (13) is replaced by the first term

$$(65) \quad \gamma(W) = \frac{1}{\nu} k_1 L_1(W) = T(1) L_1(W).$$

This follows from the approximation

$$(66) \quad L_p(W) = 2^p p! \frac{J_p(W)}{W^p} \simeq \frac{2^{p+\frac{1}{2}} p!}{\sqrt{\pi}} \frac{\cos\{W - [(2p+1)/4]\pi\}}{W^{p+\frac{1}{2}}},$$

which shows that all the functions  $L_p(W)$  can be neglected with respect to  $L_1(W)$ , whence

$$(67) \quad \begin{aligned} E(W) &= \frac{\int_0^W \gamma^2(W) d(W^2)}{\int_0^\infty \gamma^2(W) d(W^2)} = 1 - \frac{\int_W^\infty \gamma^2(W) d(W^2)}{\int_0^\infty \gamma^2(W) d(W^2)} \\ &= 1 - \frac{1}{4\tau} \int_W^\infty L_1^2(W) d(W^2) \simeq e(W) = 1 - \frac{T^2(1)}{\tau} \int_W^\infty L_1^2(W) d(W^2). \end{aligned}$$

The integral

$$\int_W^\infty L_1^2(W) d(W^2)$$

can be derived from the factor of encircled energy of an Airy pattern:

$$(68) \quad \begin{aligned} \mathcal{E}(W) &= \frac{\int_0^W L_1^2(W) d(W^2)}{\int_0^\infty L_1^2(W) d(W^2)} = 1 - \frac{\int_W^\infty L_1^2(W) d(W^2)}{\int_0^\infty L_1^2(W) d(W^2)} \\ &= 1 - \frac{1}{4} \int_W^\infty L_1^2(W) d(W^2), \end{aligned}$$

whence the approaching expression  $e(W)$  of  $E(W)$

$$(69) \quad e(W) = 1 - \frac{T^2(1)}{\tau} [1 - \mathcal{E}(W)].$$

The approximation  $e(W)$  instead of  $E(W)$  can be appreciated from Table VIII. It can be readily seen that  $e(W)$  is valid for  $W > 10$  with an accuracy of four significant digits.

Finally, the coefficient  $R = T^2(1)/\tau$ , which represents the rate of reduction of light diffracted in the remote region compared to the Airy pattern. The expression of  $R$  can be written as

$$(70) \quad R = \frac{T^2(1)}{\tau} = \frac{1-e(W)}{1-\mathcal{E}(W)},$$

which means

$$R = \frac{\text{energy in the diffraction pattern outside a circle of radius } W}{\text{energy in the Airy pattern outside a circle of radius } W}.$$

In the cases considered here, that is  $W_m = 2, 3, 4, 5$ ,  $R$  takes the respective values:  $R = 0.58, 0.26, 0.08, 0.02$ . In other words, those amplitude filters allow a separation whose distance from a very bright point is  $W$ , and whose intensity is 2, 4, 12, 50 times smaller.

Some applications can be suggested immediately as, for example, the separation of points relatively distant but having unequal intensities. These cases frequently occur in astronomy and spectroscopy.

Independently of these particular cases, a high rate of reduction in intensity of the outer part of the diffraction pattern should improve the quality of optical images. For example, in the image of an extended object this would mean reduction of fringes along the edges and hence an increased over-all sharpness of the picture.

#### X. ENVELOPE $E_n(W_m, W_m)$ OF CURVES OF FACTOR OF ENCIRCLED ENERGY $E_n(W, W_m)$

To avoid any confusion in this section the notation  $\Gamma_n(W)$  will be replaced by  $\Gamma_n(W, W_m)$  meaning that the expansion of order  $n$  with respect to  $W$  for the amplitude concerns the diffraction pattern corresponding to the maximum factor of encircled energy with  $W = W_m$ . In the same way,  $E_n(W, W_m)$  stands for the factor of encircled energy relative to  $\Gamma_n(W, W_m)$  with respect to  $W$ . Finally,  $\Gamma(W, W_m)$  and  $E(W, W_m)$  are the amplitude and factor of encircled energy when  $n \rightarrow \infty$ .

Figures 5 to 9 show the family of curves  $E(W, W_m)$  for  $W_m = 2, 3, 4, 5$  and  $0 < W < 10$ . The envelope of these curves is the locus of extrema  $E(W_m, W_m) = M(W)$ .

This property holds for any order  $n$ . In other words, the curves  $E_n(W_m, W_m)$  and  $E_n(W, W_m)$  are tangent at  $W = W_m$ .

This property can be shown from the definition of  $E_n(W, W_m)$ . We have

$$(71) \quad \int_0^W \Gamma_n^2(W, W_m) d(W^2) - E_n(W, W_m) \int_0^\infty \Gamma_n^2(W, W_m) d(W^2) = 0.$$

TABLE VIII  
Factor of encircled energy  $E(W)$  and its asymptotic expression  $e(W) = 1 - [T^2(1)/\tau][1 - \mathcal{E}(W)]$

W	$\mathcal{E}(W)$	$W_m = 0$		$W_m = 2$		$W_m = 3$		$W_m = 4$		$W_m = 5$	
		$E(W)$	$e(W)$								
0.0	0.0000	0.0000	0.4229	0.0000	0.7379	0.0000	0.9175	0.0000	0.9796	0.0000	0.9796
0.5	0.0606	0.0565	0.5504	0.0557	0.2065	0.7358	0.0493	0.9426	0.0426	0.9606	0.0426
1.0	0.2208	0.2157	0.4279	0.4100	0.6149	0.8997	0.3721	0.9357	0.3285	0.9841	0.3285
1.5	0.4267	0.4227	0.6296	0.7791	0.7802	0.8997	0.5700	0.9684	0.5129	0.9922	0.5129
2.0	0.6173	0.6173	0.7506	0.7810	0.8676	0.8871	0.7416	0.8660	0.6828	0.9963	0.6828
2.5	0.7506	0.7506	0.8174	0.8676	0.9013	0.9398	0.9403	0.9403	0.9090	0.9963	0.9090
3.0	0.8366	0.8366	0.8379	0.9071	0.9065	0.9568	0.9575	0.9750	0.9866	0.9967	0.9866
3.5	0.8439	0.8439	0.9081	0.9153	0.9199	0.9589	0.9636	0.9861	0.9866	0.9953	0.9953
4.0	0.8612	0.8612	0.9279	0.9327	0.9642	0.9599	0.9877	0.9877	0.9985	0.9972	0.9972
4.5	0.8834	0.8834	0.9398	0.9427	0.9702	0.9642	0.9879	0.9904	0.9971	0.9976	0.9976
5.0	0.9008	0.9008	0.9465	0.9473	0.9749	0.9740	0.9893	0.9918	0.9972	0.9980	0.9980
5.5	0.9087	0.9087	0.9482	0.9480	0.9765	0.9749	0.9911	0.9925	0.9975	0.9981	0.9981
6.0	0.9108	0.9108	0.9484	0.9485	0.9766	0.9766	0.9924	0.9926	0.9979	0.9982	0.9982
6.5	0.9155	0.9155	0.9504	0.9512	0.9771	0.9771	0.9928	0.9930	0.9983	0.9983	0.9983
7.0	0.9236	0.9236	0.9548	0.9559	0.9787	0.9800	0.9930	0.9937	0.9983	0.9984	0.9984
7.5	0.9317	0.9317	0.9587	0.9606	0.9810	0.9821	0.9937	0.9944	0.9984	0.9986	0.9986
8.0	0.9364	0.9364	0.9629	0.9633	0.9828	0.9833	0.9944	0.9948	0.9985	0.9987	0.9987
8.5	0.9376	0.9376	0.9640	0.9640	0.9836	0.9837	0.9948	0.9948	0.9987	0.9987	0.9987

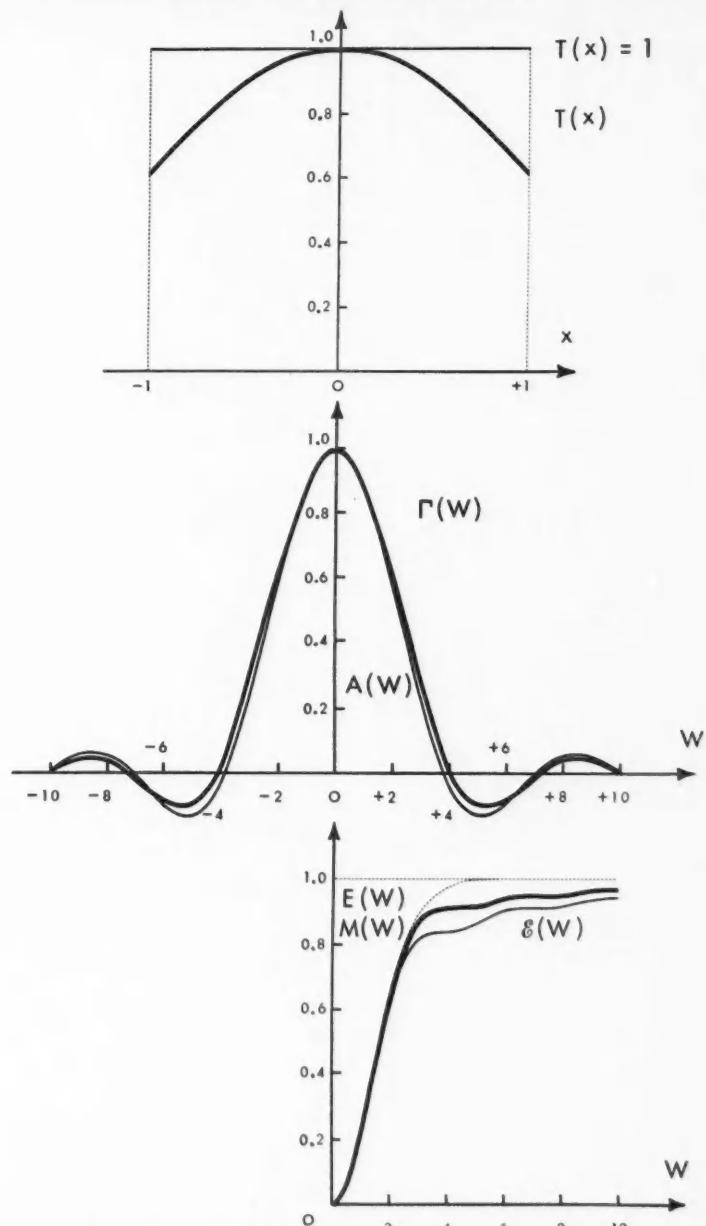


FIG. 5. Curves of the limit functions  $T(x)$ ,  $\Gamma(W)$ ,  $E(W)$  respectively compared to the uniform transparency system  $T(x) = 1$ , to the Airy pattern  $A(W)$ , and the corresponding factor of encircled energy  $\mathcal{E}(W)$  for  $W_m = 2$ . The envelope of the maxima of  $E(W_m)$  is the dotted line  $M(W)$ .

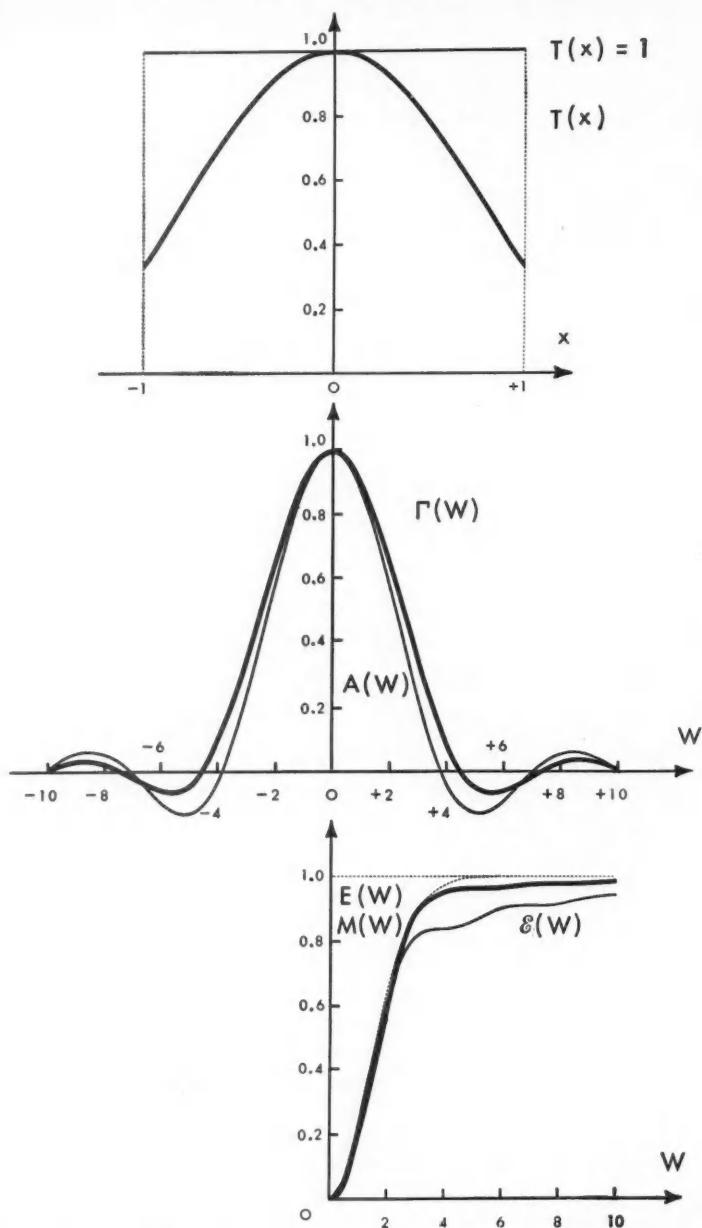


FIG. 6. Curves of the limit functions  $T(x)$ ,  $\Gamma(W)$ ,  $E(W)$  respectively compared to the uniform transparency system  $T(x)=1$ , to the Airy pattern  $A(W)$ , and the corresponding factor of encircled energy  $\mathcal{E}(W)$  for  $W_m = 3$ . The envelope of the maxima of  $E(W_m)$  is the dotted line  $M(W)$ .

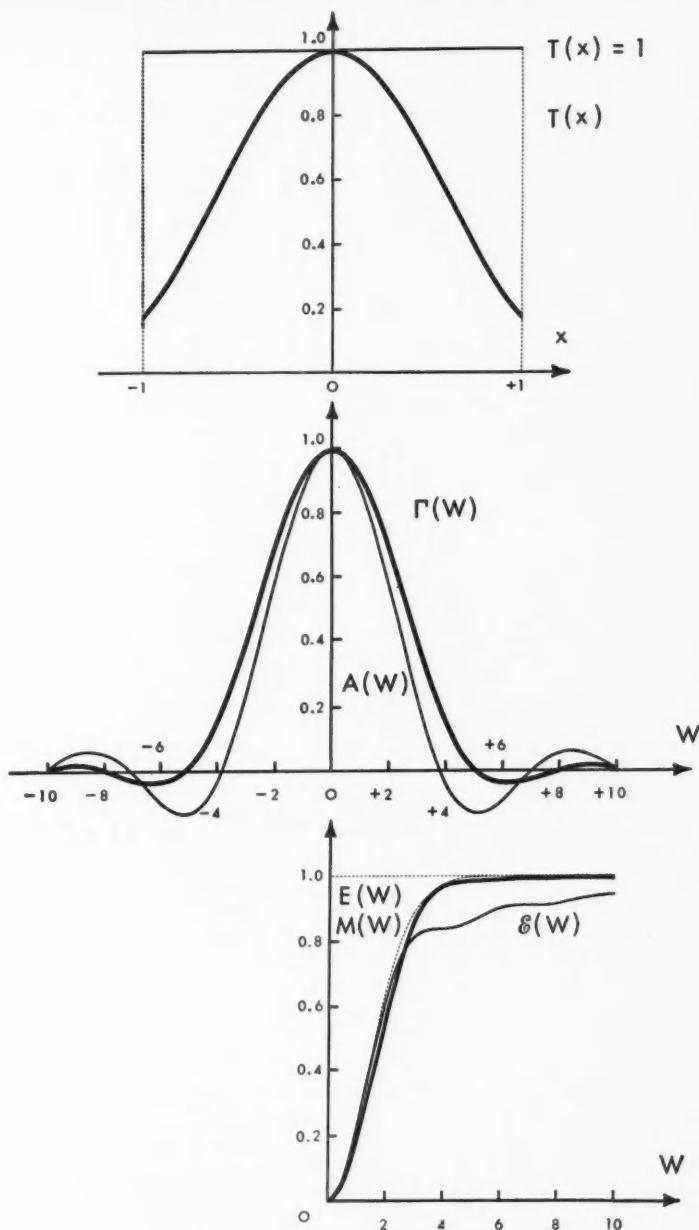


FIG. 7. Curves of the limit functions  $T(x)$ ,  $\Gamma(W)$ ,  $E(W)$  respectively compared to the uniform transparency system  $T(x) = 1$ , to the Airy pattern  $A(W)$ , and the corresponding factor of encircled energy  $\mathcal{E}(W)$  for  $W_m = 4$ . The envelope of the maxima of  $E(W_m)$  is the dotted line  $M(W)$ .

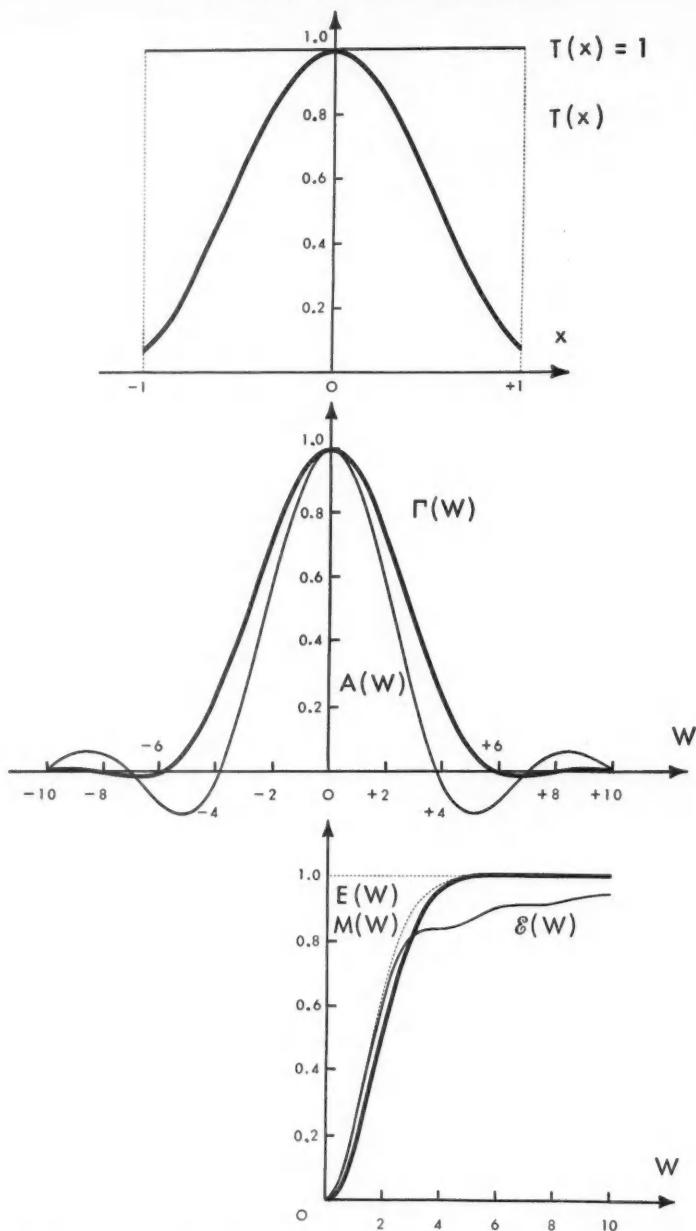
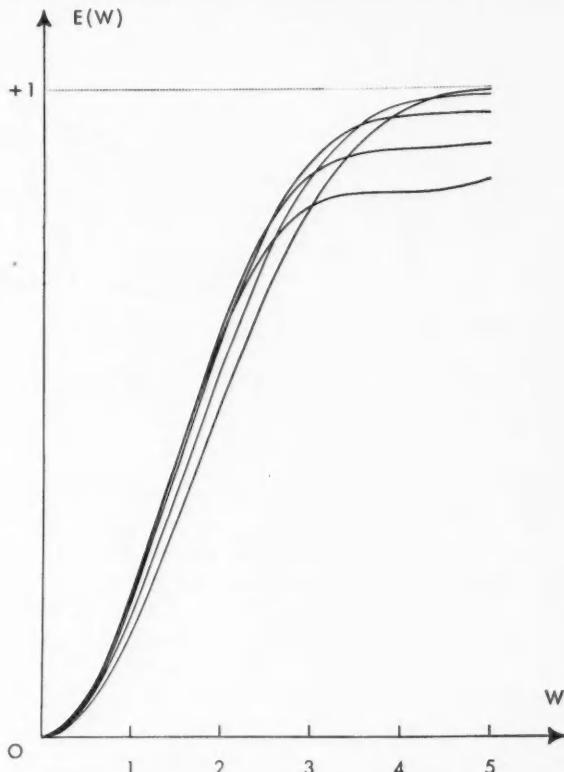


FIG. 8. Curves of the limit functions  $T(x)$ ,  $\Gamma(W)$ ,  $E(W)$  respectively compared to the uniform transparency system  $T(x) = 1$ , to the Airy pattern  $A(W)$ , and the corresponding factor of encircled energy  $\mathcal{E}(W)$  for  $W_m = 5$ . The envelope of the maxima of  $E(W_m)$  is the dotted line  $M(W)$ .

FIG. 9. Family of curves  $E(W)$  for  $W_m = 0, 2, 3, 4, 5$ .

For a fixed value of  $W_m$  the partial derivative  $[\partial E_n(W, W_m)/\partial W]_{W=W_m}$  is the slope of the tangent to the curve of the factor of encircled energy  $E_n(W, W_m)$  for  $W = W_m$ . The derivatives of (71) with respect to  $W$  at  $W = W_m$  is

$$(72) \quad 2W_m \Gamma_n^2(W_m, W_m) - \left[ \frac{\partial E_n(W, W_m)}{\partial W} \right]_{W=W_m} \int_0^\infty \Gamma_n^2(W, W_m) d(W^2) = 0.$$

To find the slope  $dE_n(W_m, W_m)/dW_m$  of the tangent to the curve  $E_n(W_m, W_m)$  we shall differentiate (71) with respect to  $W_m$  setting  $W = W_m$ , that is,

$$(73) \quad \int_0^{W_m} \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) - E_n(W_m, W_m) \int_0^\infty \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) \\ + 2W_m^2 \Gamma_n^2(W_m, W_m) - \frac{dE_n(W_m, W_m)}{dW_m} \int_0^\infty \Gamma_n^2(W, W_m) d(W^2) = 0.$$

But

$$\begin{aligned}
 \int_0^{W_m} \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) &= 2 \int_0^{W_m} \Gamma_n(W, W_m) \frac{\partial \Gamma_n(W, W_m)}{\partial W_m} d(W^2) \\
 &= 2 \sum_{p=1}^{p=n} \frac{dk_p^n}{dW_m} \int_0^{W_m} \Gamma_n(W, W_m) L_p(W) d(W^2), \\
 (74) \quad \int_0^{\infty} \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) &= 2 \int_0^{\infty} \Gamma_n(W, W_m) \frac{\partial \Gamma_n(W, W_m)}{\partial W_m} d(W^2) \\
 &= 2 \sum_{p=1}^{p=n} \frac{dk_p^n}{dW_m} \int_0^{\infty} \Gamma_n(W, W_m) L_p(W) d(W^2).
 \end{aligned}$$

From the system of equation (19) it can be written

$$\begin{aligned}
 (75) \quad \int_0^{W_m} \Gamma_n(W, W_m) d(W^2) - E_n(W_m, W_m) \int_0^{\infty} \Gamma_n(W, W_m) d(W^2) \\
 \quad \text{with } p = 1, \dots, n
 \end{aligned}$$

and one deduces by linear combination that the expression (74) satisfies:

$$(76) \quad \int_0^{W_m} \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) - E_n(W_m, W_m) \int_0^{\infty} \frac{\partial \Gamma_n^2(W, W_m)}{\partial W_m} d(W^2) = 0,$$

whence the expression (73) is reduced to

$$(77) \quad 2W_m \Gamma_n^2(W_m, W_m) - \frac{dE(W_m, W_m)}{dW_m} \int_0^{\infty} \Gamma_n^2(W, W_m) d(W^2) = 0.$$

Comparing this relation with (72) it is seen that

$$(78) \quad \left[ \frac{dE_n(W_m, W_m)}{dW_m} = \frac{\partial E_n(W, W_m)}{\partial W} \right]_{W=W_m},$$

which proves that the curves  $E_n(W_m, W_m)$  and  $E(W, W_m)$  are tangent at  $W = W_m$ .

The property, being independent of  $n$ , holds also for  $n \rightarrow \infty$ .

The envelope  $E(W_m, W_m)$  has an important physical meaning: it is the limit of the factor of encircled energy or, the limit of performance of optical instruments.

#### CONCLUSION

The formulae, the numerical calculations, and the graphs presented in this paper show how amplitude filters can improve the image forming quality of optical systems. The improvement results from a concentration of energy in the central part of the diffraction image of a point source, and hence a consequent reduction in background intensity. This helps a better reproduction in the image, the contrast of extended objects also.

These theoretical results are now being subjected to experimental verification. Initial results are found to be encouraging. It is proposed to report the final results in due course.

It is interesting to note that there is a maximum limit to the concentration of energy in the central core of the diffraction pattern, and the locus  $E(W, W_m)$  as given by the envelope  $M(W)$  represents the limiting performance of optical systems.

#### SOMMAIRE

On admet que la plus favorable des répartitions d'énergie lumineuse dans la figure de diffraction est celle qui correspond au maximum de concentration autour de son centre  $O$ . Cette hypothèse est exprimée par une condition extrême sur le facteur d'énergie encerclée  $E(W)$ , c'est-à-dire sur le rapport de l'énergie contenue dans le cercle de rayon  $W$ , centré en  $O$ , à l'énergie totale de la figure de diffraction.

L'étude de l'aberration de sphéricité montre que le maximum absolu du facteur d'énergie encerclée  $E(W_m)$  pour un rayon donné  $W = W_m$ , n'est atteint qu'avec un système optique dénué d'aberration, correctement mis au point et dont la transmission  $T(x)$  est rendue localement variable par un filtre d'amplitude.

On représente la transmission  $T(x)$ , soit de façon rigoureuse par la série de Taylor en  $(1-x^2)^n$ , soit de façon approchée par un polynôme  $T_n(x)$  de degré  $n-1$  en  $1-x^2$ .

Les amplitudes correspondantes dans la figure de diffraction sont  $\Gamma(W)$  et  $\Gamma_n(W)$  et les facteurs d'énergie encerclée  $E(W_m)$  et  $E_n(W_m)$ . On établit les convergences  $E_n(W_m) \rightarrow E(W_m)$ ,  $T_n(x) \rightarrow T(x)$  et  $\Gamma_n(W) \rightarrow \Gamma(W)$ , lorsque  $n \rightarrow \infty$ .

Lorsque la normalisation des abscisses  $W$  et  $x$ , fait correspondre le domaine  $(0, W_m)$  de la figure de diffraction au domaine  $(0, 1)$  de la pupille, les distributions d'amplitudes  $\Gamma(W)$  et  $T(x)$  sont identiques; soit:  $T(x) = \Gamma(W_m x) = \Gamma(W)$  avec  $W = W_m x$ .

De cette relation sont déduites différentes formules ou propriétés concernant  $T(x)$  et  $\Gamma(W)$ . Ainsi la tache d'Airy est la limite de  $\Gamma(W)$  lorsque  $W_m \rightarrow 0$ . Par contre, lorsque  $W_m$  croît indéfiniment, la fonction de Gauss  $e^{-(W_m/2)x^2}$  est une représentation asymptotique de  $T(x)$ . Dans tous les cas, le maximum du facteur d'énergie encerclée  $E(W_m)$  est lié à la transmission au bord de la pupille par  $E(W_m) = 1 - T^2(1)$ .

Les déterminations numériques de  $E(W)$ , faites ici pour  $0 < W < 10$  et  $W_m = 2, 3, 4$  et  $5$ , peuvent être complétées aux grandes valeurs de  $W$  grâce à une expression asymptotique du facteur d'énergie encerclé.

Les courbes  $E(W)$  correspondant à toutes les valeurs  $W_m$  admettent une enveloppe  $M(W)$ , lieu des maxima du facteur d'énergie encerclé. Cette enveloppe constitue la limite de performance des instruments d'optique.

#### ACKNOWLEDGMENTS

The author gratefully acknowledges financial assistance through a research grant from the National Research Council and a studentship to one of us (G.B.). We are also indebted to Dr. L. Gauvin and Dr. M. De for helpful discussions on the original manuscript.

## REFERENCES

DOSSIER, B. 1952. *J. phys. radium*, **13**, 169.  
——— 1954. Thèse.  
LANSRAUX, G. 1946. *Compt. rend.* **222**, 1434.  
——— 1947. *Rev. opt.* **26**, 24.  
——— 1952. *Rev. opt.* **31**, 321.  
——— 1953. *Rev. opt.* **32**, 475.  
LANSRAUX, G. and BOIVIN, G. 1958. *Can. J. Phys.* **36**, 1696.  
OSTERBERG, H. and WILKINS, J. E. 1949. *J. Opt. Soc. Am.* **39**, 553.  
STRAUBEL, R. 1935. *Zeeman Verhandelingen* (M. Nijhoff), 302.  
WOLF, E. 1951. *Repts. Progr. in Phys.* **14**, 95.

# THEORY OF TRANSLATIONAL ABSORPTION IN GASES<sup>1</sup>

J. D. POLL<sup>2</sup> AND J. VAN KRANENDONK

## ABSTRACT

The theory of translational infrared absorption in gases is developed. Invariant expressions for the integrated absorption coefficients are derived. The absorption coefficients are expanded in powers of the density, and the binary absorption coefficients are expressed in terms of a model for the induced pair dipole moments. Monatomic gas mixtures, diatomic gases, and diatomic-monatomic gas mixtures are considered in detail. As an application the binary absorption coefficient of the translational band of hydrogen is calculated.

## 1. INTRODUCTION

The pressure-induced infrared spectra of gases arise from the dipole moments induced by the intermolecular forces in clusters of two or more interacting molecules. These induced dipole moments decrease rapidly in magnitude with increasing intermolecular separation, and vanish beyond two or three molecular diameters. This property of the induced dipole moments of being short-range functions of the intermolecular separations constitutes the most characteristic difference between pressure-induced infrared activity on the one hand and ordinary infrared activity on the other.

The induced rotational and vibrational absorption bands are due to the modulation of the induced dipole moments by the rotational and vibrational motions of the molecules. Because of the strong dependence of the induced dipole moments on the intermolecular separations, the dipole moments are also modulated by the translational motion of the molecules. The translational motion is therefore infrared active, and plays an important role in all pressure-induced spectra. In the absorption processes responsible for the induced rotational and vibrational spectra of hydrogen, for example, the rotational and vibrational transitions are accompanied by transitions between different translational states of the molecules. This "translational" effect is responsible for the width of the induced infrared lines (Crawford, Welsh, MacDonald, and Locke 1950; Van Kranendonk 1957), for the characteristic shape of the *Q*-branch of the fundamental band of hydrogen (Chisholm and Welsh 1954), and for the appearance of the phonon branches in the induced spectra of solids (Gush, Hare, Allin, and Welsh 1960).

It is clear that in pressure-induced absorption pure translational transitions not accompanied by any rotational or vibrational transitions are also possible. The resulting translational absorption band lies in the far infrared frequency region, and extends from zero frequency up to frequencies of the order of magnitude of the rotational frequencies. Experimental evidence for such a

<sup>1</sup>Manuscript received October 6, 1960.

Contribution from the Department of Physics, University of Toronto, Toronto, Ontario. This research was supported in part by a grant from the National Research Council of Canada.

<sup>2</sup>Holder of a Graduate Fellowship in the Computation Center of the University of Toronto, 1958-1959, and a University of Toronto Open Fellowship, 1959-1960.

translational band has been obtained by Kiss and Welsh (1959b) in the rotational spectrum of hydrogen, both in the pure gas and in mixtures with foreign gases. The translational absorption band has also been observed in mixtures of the rare gases (Kiss and Welsh 1959a).

In the present paper general theoretical expressions are derived for the integrated intensity of the translational absorption bands in monatomic and diatomic gases. A comparison of these expressions with the experimental data on the integrated intensities yields information on the nature of the induced dipole moments. For example, the temperature dependence of the binary absorption coefficient (Van Kranendonk 1958) is determined by the nature of the dependence of the dipole moment induced in a pair of molecules on the intermolecular separation, whereas the intensity of the absorption is determined by the magnitude of the induced dipole moments. Ideally, measurements of the binary absorption coefficient as a function of the temperature make possible an empirical determination of the induced pair dipole moment, in much the same way as a knowledge of the second virial coefficient of a gas allows a determination of the potential of the intermolecular forces.

In a monatomic gas mixture only the translational band appears, and no other infrared absorption bands are present. In a homonuclear diatomic gas, on the other hand, there appear a translational and a rotational band, as well as a vibrational band. The translational and rotational bands are in general not separated from each other, and form essentially one band which we call the far infrared band. We assume that this far infrared band is completely separated from the vibrational bands, as is usually the case. The total infrared intensity is then equal to the sum of the intensities of the vibrational and far infrared bands. A similar separation of the translational and rotational bands is not possible, and in general the far infrared band does not consist additively of a translational and a rotational part. Such a separation is possible only if there are no anisotropic intermolecular forces. In the presence of anisotropic intermolecular forces, the translational and rotational motions of the molecules are coupled, and no rigorous separation of the far infrared band into two separate parts can be made.

In Section 2 general expressions for the integrated absorption coefficients of the far infrared spectrum are derived, and in Section 3 the expansion of these quantities in powers of the density is given. In Section 4 the binary absorption coefficient of a monatomic gas mixture is expressed in terms of the pair dipole moment, and similarly in Section 5 for a diatomic gas and for a diatomic-monatomic gas mixture. In Section 6 the  $(\exp - 4)$  model (Van Kranendonk 1958; Van Kranendonk and Kiss 1959) for the pair dipole moment is introduced, and the binary absorption coefficient for hydrogen is calculated.

## 2. INVARIANT EXPRESSIONS FOR THE INTEGRATED ABSORPTION COEFFICIENTS

The general expression for the integrated absorption coefficient of the far infrared spectrum,

$$(1) \quad \alpha = \int A(\nu) d\nu,$$

where  $A(\nu)$  is the absorption coefficient per unit path length and  $\nu$  is the frequency, is given by

$$(2) \quad \alpha = \kappa V^{-1} \sum_{i,f} (P_i - P_f) |\psi_{if}|^2 \hbar \nu_{if}.$$

Here  $\kappa = (8\pi^3/3h^2c)$ , and  $V$  is the total volume of the medium. The subscripts  $i$  and  $f$  denote stationary states of the system. For monatomic gases, these states are the translational states of the molecules, and for diatomic gases the translation-rotational states. We assume that the molecules are permanently in the ground vibrational and electronic state. The dipole moment  $\psi$  appearing in (2) is then equal to the expectation value of the total dipole moment operator of the gas over the ground vibrational and electronic state. This dipole moment is a function of the nuclear configuration of the gas, and  $\psi_{if}$  denotes the matrix element of  $\psi$  between the states  $i$  and  $f$ . The quantities  $P_i$  and  $P_f$  appearing in (2) are the normalized Boltzmann factors for the states  $i$  and  $f$ , and  $\hbar \nu_{if} = E_f - E_i$ , where  $E_i$  and  $E_f$  are the energies of the states  $i$  and  $f$  respectively. Finally, the sum in (2) runs over all pairs of states  $i, f$  for which  $E_i < E_f$ . The terms containing  $P_i$  and  $P_f$  are due to the absorption and stimulated emission processes respectively.

The expression (2) can be written in the form of a trace over a complete set of states by making use of the fact that the summand in (2) is symmetric in  $i$  and  $f$ , and of the Hermitian property of  $\psi_{if}$ . In virtue of these properties, the expression (2) is clearly equal to

$$(3) \quad \alpha = \kappa V^{-1} \sum_{i,f} P_i |\psi_{if}|^2 \hbar \nu_{if},$$

where the sum now extends over all values of  $i$  and  $f$ . The expression (3) can be written in the form of the invariant trace,

$$(4) \quad \alpha = \kappa V^{-1} \text{Tr}\{P \psi \cdot [H, \psi]\},$$

where  $H$  is the Hamiltonian of the translational, or translation-rotational, motion of the molecules;  $[H, \psi] = H\psi - \psi H$  denotes the commutator; and  $P$  is the density of the ensemble representing the system in equilibrium at a given temperature  $T = 1/k\beta$ ,

$$(5) \quad P = \exp(-\beta H) / \text{Tr}[\exp(-\beta H)].$$

The invariant expression (4) forms the basis of the calculation of the integrated intensities of the far infrared spectra.

In the theory of the induced vibrational (Van Kranendonk 1958) and rotational (Van Kranendonk and Kiss 1959) spectra, on the other hand, the integrated absorption coefficient  $\bar{\alpha}$ , defined as

$$(6) \quad \bar{\alpha} = \int \bar{A}(\nu) d\nu,$$

where  $\tilde{A}(\nu) = A(\nu)\lambda$  is the absorption coefficient per wavelength *in vacuo*, is used instead of  $\alpha$ . The general expression for  $\tilde{\alpha}$  is

$$(7) \quad \tilde{\alpha} = \tilde{\kappa} V^{-1} \sum_{i < f} (P_i - P_f) |\mathbf{u}_{if}|^2,$$

where  $\tilde{\kappa} = (8\pi^3/3h)$ . From a theoretical point of view, the quantity  $\tilde{\alpha}$  is more fundamental than the quantity  $\alpha$ , because  $\tilde{\alpha}$  is proportional to the total transition probability, whereas  $\alpha$  is proportional to the absorbed energy, provided the incident intensity is independent of the frequency. Mathematically,  $\tilde{\alpha}$  and  $\alpha$  are proportional to the zeroth and first moment respectively of the function  $\tilde{A}(\nu)$ .

For a spectrum extending to zero frequency, however, the quantity  $\tilde{\alpha}$  does not lead to convenient expressions, and, as we shall show, one must therefore use  $\alpha$  instead. The expression (7) for  $\tilde{\alpha}$  can be written in the form of a trace by multiplying the summand in (7) by the unit step function

$$(8) \quad \epsilon(\nu_{if}) = (1/i\pi) \int_{-\infty}^{+\infty} t^{-1} \exp(2\pi i \nu_{if} t) dt,$$

where the principal value of the integral should be taken. In this way we get the following invariant expression for  $\tilde{\alpha}$ ,

$$(9) \quad \tilde{\alpha} = (\tilde{\kappa}/i\pi V) \int_{-\infty}^{+\infty} t^{-1} \text{Tr}[P \mathbf{u}(0) \cdot \mathbf{u}(t)] dt,$$

where  $\mathbf{u}(t)$  is the Heisenberg time-dependent dipole moment. The expression (9) for  $\tilde{\alpha}$  is much less useful than the expression (4) for  $\alpha$ , because the complete time dependence of the dipole moment is involved in (9). However, if the spectrum does not extend to zero frequency, the expression (7) for  $\tilde{\alpha}$  can be reduced to a much simpler expression than the general expression (9). As an example, consider the induced vibrational spectrum of hydrogen (cf. Van Kranendonk 1958). In this case, the vibrational excitation energy is much larger than the changes in the rotational and translational energy occurring during the absorption processes. This means that the vibrational absorption band, the structure of which is determined by the rotational and translational transitions accompanying the vibrational transition, does not extend to zero frequency. The restrictive condition  $E_i < E_f$ , appearing in (7) is then satisfied by all transitions contributing significantly to the total intensity, and the sum in (7) may then be extended over all translation-rotational states  $i$  and  $f$ . When we denote the matrix element of the dipole moment between the relevant initial and final vibrational state by  $\mathbf{M}$ , the total integrated absorption coefficient  $\tilde{\alpha}$  of the corresponding vibrational band is given by

$$(10) \quad \tilde{\alpha} = \tilde{\kappa} \text{Tr}[P |\mathbf{M}|^2],$$

where  $\text{Tr}$  indicates a trace over the translation-rotational states, and  $P$  is the density for these states, provided all the molecules are originally in the initial vibrational state. The expression (10) for  $\tilde{\alpha}$  has been used in the theory

of the induced vibrational and rotational spectra of hydrogen (Van Kranendonk 1957, 1958; Van Kranendonk and Kiss 1959), since it is much simpler than the expression (4) for  $\alpha$ .

For the translational spectrum, however, the expression (10) for  $\tilde{\alpha}$  cannot be used, since the spectrum extends to zero frequency. The translational spectrum can be interpreted as a line at zero frequency broadened by the translational effect, and an expression of the type (10) is not even approximately true in this case. One must therefore use either the expression (9) for  $\tilde{\alpha}$ , or the expression (4) for  $\alpha$ . In the present paper, we use exclusively the simpler expression (4). It is also possible to derive invariant expressions for the higher moments of the spectrum, like  $\int A(\nu) \nu^2 d\nu$ , but we shall not discuss these higher moments in the present paper.

### 3. DENSITY EXPANSION OF THE INTEGRATED ABSORPTION COEFFICIENTS

In this section the integrated absorption coefficient (4) of the far infrared spectrum of gases is expanded in powers of the density. As in the case of the vibrational (Van Kranendonk 1958) and rotational (Van Kranendonk and Kiss 1959) spectra, this expansion can be obtained by introducing appropriate cluster functions. The expansion is possible in virtue of the fact that the induced dipole moments are short-range functions of the intermolecular separations, i.e. decrease faster than  $R^{-3}$  with increasing separation  $R$ .

We first consider a single component diatomic gas containing  $N$  molecules, with a total volume  $V$ . Let  $\mathbf{y}(1 \dots n)$  be the dipole moment induced in the cluster of molecules  $1, \dots, n$  when these are present alone in the volume  $V$ . We introduce the cluster functions  $\mathbf{U}(1 \dots n)$  by means of the equations

$$(11) \quad \left\{ \begin{array}{l} \mathbf{y}(12) = \mathbf{U}(12), \\ \mathbf{y}(123) = \mathbf{U}(12) + \mathbf{U}(13) + \mathbf{U}(23) + \mathbf{U}(123), \\ \dots \\ \mathbf{y}(1 \dots N) = \sum_{i < j} \mathbf{U}(ij) + \sum_{i < j < k} \mathbf{U}(ijk) + \dots \end{array} \right.$$

These equations can be solved successively for the  $\mathbf{U}$ 's in terms of the  $\mathbf{y}$ 's. When these expressions are substituted into the last of the equations (11), this equation reduces to an identity which then represents the cluster development of  $\mathbf{y}(1 \dots N)$ . The symbols  $1, 2, \dots$ , in the arguments of the functions  $\mathbf{y}$  and  $\mathbf{U}$  denote the nuclear co-ordinates of the molecules, i.e.  $i \equiv (\mathbf{R}_i, \omega_i)$ , where  $\mathbf{R}_i$  is the position vector of the center of mass, and  $\omega_i \equiv (\theta_i, \phi_i)$  specifies the orientation of the internuclear axis of molecule  $i$ . Since  $\mathbf{y}(1 \dots N)$  is a function of the nuclear co-ordinates only,  $\mathbf{y}$  commutes with the potential energy term in the Hamiltonian  $H$ , and we may therefore replace  $H$  in the expression (4) for  $\alpha$  by the kinetic energy

$$(12) \quad K = \sum_i K(i),$$

where  $K(i)$  is the kinetic energy of the translational and rotational motion of molecule  $i$ .

The expansion of  $\alpha$  in powers of the density  $n = N/V$ ,

$$(13) \quad \alpha = \alpha_1 n^2 + \alpha_2 n^3 + \dots,$$

is obtained by substituting the last of the equations (11), and the expression (12), into (4). The binary absorption coefficient  $\alpha_1$  is given by

$$(14) \quad \alpha_1 = \frac{1}{2} \kappa V^{-1} \text{Tr}[P_2^{(0)}(12) \mathbf{y}(12) \cdot \hat{\mathbf{y}}(12)],$$

where  $P_2^{(0)}(12)$  is the low density limit of the pair density operator  $P_2(12)$  (De Boer 1948),

$$(15) \quad P_2(12) = P_2^{(0)}(12) + P_2^{(1)}(12)n + \dots.$$

The quantity  $\hat{\mathbf{y}}(12)$  in (14) denotes the commutator

$$(16) \quad \hat{\mathbf{y}}(12) \equiv [K, \mathbf{y}(12)] = [K(1) + K(2), \mathbf{y}(12)].$$

Similar definitions hold for  $\hat{\mathbf{y}}(1 \dots n)$  and  $\hat{\mathbf{U}}(1 \dots n)$ . The ternary absorption coefficient  $\alpha_2$  appearing in (13) is equal to a sum of three terms,

$$(17) \quad \alpha_2 = \alpha_2^{(1)} + \alpha_2^{(2)} + \alpha_2^{(3)},$$

where

$$(18) \quad \alpha_2^{(1)} = \frac{1}{2} \kappa V^{-1} \text{Tr}[P_2^{(1)}(12) \mathbf{y}(12) \cdot \hat{\mathbf{y}}(12)]$$

is the contribution due to the density dependence of  $P_2$ ;

$$(19) \quad \alpha_2^{(2)} = \kappa V^{-1} \text{Tr}[P_3^{(0)}(123) \mathbf{y}(12) \cdot \hat{\mathbf{y}}(13)]$$

represents an interference effect between  $\mathbf{y}(12)$  and  $\mathbf{y}(13)$ , and

$$(20) \quad \alpha_2^{(3)} = \frac{1}{2} \kappa V^{-1} \text{Tr} \left[ P_3^{(0)}(123) \left\{ \mathbf{y}(12) \cdot \hat{\mathbf{U}}(123) + \mathbf{U}(123) \cdot \hat{\mathbf{y}}(12) \right. \right. \\ \left. \left. + \frac{1}{3} \mathbf{U}(123) \cdot \hat{\mathbf{U}}(123) \right\} \right]$$

is the contribution from the non-additive part

$$\mathbf{U}(123) = \mathbf{y}(123) - \mathbf{y}(12) - \mathbf{y}(13) - \mathbf{y}(23)$$

of the dipole moment induced in a cluster of three molecules.

For a mixture of two gases,  $a$  and  $b$ , the corresponding expressions for the absorption coefficients can be easily derived. For a mixture of two monatomic gases, the expansion (13) takes the form

$$(21) \quad \alpha = \alpha_1 n_a n_b + \alpha_{2a} n_a^3 + \alpha_{2b} n_b^3 + \alpha'_{2a} n_a^2 n_b + \alpha'_{2b} n_a n_b^2 + \dots$$

The binary absorption coefficient  $\alpha_1$  is given by

$$(22) \quad \alpha_1 = \kappa V^{-1} \text{Tr}[P_2^{(0)}(11') \mathbf{y}(11') \cdot \hat{\mathbf{y}}(11')],$$

where 1 and 1' refer to molecules of the type  $a$  and  $b$  respectively, and  $1 \equiv \mathbf{R}_1$  denotes the center of mass co-ordinates of molecule 1. We shall not write down the explicit expressions for the ternary absorption coefficients. Finally, for a mixture of a diatomic and a monatomic gas we get

$$(23) \quad \alpha = \alpha_1 n_a^2 + \alpha'_1 n_a n_b + \dots,$$

where  $\alpha_1$  is given by (14), and  $\alpha'_1$  by (22), where  $1 \equiv (\mathbf{R}_1, \omega_1)$  and  $1' \equiv \mathbf{R}'_1$ .

The binary absorption coefficient (22) for a monatomic gas mixture is worked out further in the next section. The case of a diatomic gas and of a diatomic-monatomic gas mixture is taken up in Section 5.

#### 4. THE BINARY ABSORPTION COEFFICIENT OF MONATOMIC GAS MIXTURES

The binary absorption coefficient,  $\alpha_1(T)$ , of the translational spectrum of a mixture of two monatomic gases, which is given by the expression (22), is a function of the temperature of the gas only. An expression for this function in terms of the induced pair dipole moment will now be derived. The temperature dependence of  $\alpha_1$  will be evaluated explicitly by introducing a definite model for the pair dipole moment and for the intermolecular potential.

The trace appearing in the expression (22) for  $\alpha_1$  involves a trace over the motion of the center of mass of the two molecules 1 and 1', which can be evaluated immediately, giving

$$(24) \quad \alpha_1 = \kappa \lambda^3 \text{Tr}\{\exp(-\beta H) \mathbf{u} \cdot [K, \mathbf{u}]\},$$

where  $H$  and  $K$  denote the total and kinetic energy respectively of the relative motion of the pair of molecules, and the trace also refers to the relative motion only;  $\lambda = h/(2\pi mkT)^{1/2}$ , where  $m$  is the reduced mass of the pair of molecules. The pair dipole moment  $\mathbf{u}$  lies along the intermolecular axis  $\mathbf{R}$  and can be written in the form

$$(25) \quad \mathbf{u} = \mu(R)(\mathbf{R}/R).$$

Using the commutation relations

$$(26) \quad \begin{cases} [\Delta, f(R)] = f'' + (2/R)f' + 2f'(\partial/\partial R), \\ [\Delta, \mathbf{R}] = 2(\partial/\partial \mathbf{R}), \end{cases}$$

where the primes denote differentiation with respect to  $R$ , we get

$$(27) \quad \mathbf{u} \cdot [K, \mathbf{u}] = -(\hbar^2/2m)[\mu\mu'' + (2/R)\mu\mu' - (2/R^2)\mu^2 + 2\mu\mu'(\partial/\partial R)].$$

We write out the trace appearing in (24) in terms of the eigenfunctions

$$(28) \quad \psi_{nlm} = \chi_{nl}(R) Y_{lm}(\Omega),$$

of the Hamiltonian,  $H$ , of the relative motion of the two molecules. We use the notation  $\mathbf{R} = (R, \Omega)$ , where  $\Omega \equiv (\theta, \phi)$  specifies the orientation of  $\mathbf{R}$  relative to a space fixed co-ordinate system. The wave functions are assumed to be normalized to unity in a large spherical volume in  $\mathbf{R}$ -space with radius  $R_0$ . The low density limit of the pair distribution function normalized to unity at large separations, which we denote by  $g(R)$ , is then given by (Uhlenbeck and Beth 1936)

$$(29) \quad g(R) = \lim_{R_0 \rightarrow \infty} \lambda^3 \sum_{nl} \frac{2l+1}{4\pi} \exp(-\beta E_{nl}) |\chi_{nl}(R)|^2,$$

where  $E_{nl}$  is the energy of the state (28), and the sum extends over all the energy levels, i.e. over the bound states and over the quasi-continuous spec-

trum at positive energies. In the absence of magnetic interactions, the radial functions  $x_{nl}$  can be chosen to be real. The term arising from the last term in (27) can then be integrated by parts, and the resulting expression for the binary absorption coefficient is

$$(30) \quad \alpha_1 = (4\pi^2/3mc) \int_0^\infty [\mu'^2 + (2/R^2)\mu^2]g(R)R^2dR.$$

The first term gives the contribution to  $\alpha_1$  arising from the modulation of the dipole moment due to the variation of  $R$  during a collision, whereas the second term represents the contribution from the modulation of the dipole moment due to the variation of  $\Omega$ , i.e. of the direction of  $\mathbf{R}$ , during a collision. These two effects will be referred to as the  $R$ -modulation and  $\Omega$ -modulation effects respectively. This interpretation of the two terms in (30) can be verified by splitting up the kinetic energy  $K$  in (24) into a radial and an angle-dependent part. The resulting terms in  $\alpha_1$  are then equal to the two terms given in (30). It is also possible to derive an expression for  $\alpha_1$ , in which no distinction is made between the two modulation effects, by using Cartesian co-ordinates throughout. In this way one arrives at the expression

$$(31) \quad \alpha_1 = (\pi/3mc) \int |\nabla \mathbf{u}|^2 g(R) d\mathbf{R},$$

where

$$(32) \quad |\nabla \mathbf{u}|^2 = \sum_{ij} (\nabla_i \mu_j) * (\nabla_j \mu_i),$$

with  $\nabla_i = (\partial/\partial x_i)$  and  $d\mathbf{R} = dx_1 dx_2 dx_3$ . The two expressions (30) and (31) are easily shown to be equivalent.

The expression (30) is the desired expression for  $\alpha_1(T)$  in terms of the induced pair dipole moment. To evaluate this expression explicitly, we introduce a model for the dipole moment and for the intermolecular potential. The dipole moment induced in a pair of monatomic molecules is due to the overlap induction effect (Van Kranendonk 1958), and can be expected to depend exponentially on the intermolecular separation  $R$ . We therefore assume that  $\mu(R)$  is given by

$$(33) \quad \mu(R) = \xi \exp(-R/\rho).$$

The amplitude  $\xi$  and the range  $\rho$  are the two parameters characterizing the induced dipole moment. For the intermolecular potential we use the Lennard-Jones model,

$$(34) \quad V(R) = 4\epsilon[(\sigma/R)^{12} - (\sigma/R)^6].$$

Substituting (33) and (34) into (30), we get

$$(35) \quad \alpha_1 = \lambda^2 I_1 \gamma,$$

where  $\gamma$  is a quantity of the same dimension as  $\alpha_1$ , defined as

$$(36) \quad \gamma = (\pi e^2 \sigma^3 / 3mc),$$

$e$  being the elementary charge;  $\lambda$  is the dimensionless parameter

$$(37) \quad \lambda = (\xi/e\sigma) \exp(-\sigma/\rho),$$

and  $I_1 = I_1(T)$  is the integral

$$(38) \quad I_1 = 4\pi \int_0^\infty \exp\left[-\frac{2\sigma}{\rho}(x-1)\right] \left[\left(\frac{\sigma}{\rho}\right)^2 + \frac{2}{x^2}\right] g(x) x^2 dx,$$

where  $x = R/\sigma$ . The main contribution to the integral (38) comes from a small region around  $x = 1$ , and the contribution of the  $R$ -modulation effect is therefore roughly a factor  $\frac{1}{2}(\sigma/\rho)^2$  larger than the contribution of the  $\Omega$ -modulation effect. For the induced vibrational and rotational dipole moments in hydrogen,  $\rho/\sigma$  is equal to 0.13 (Van Kranendonk and Kiss 1959). The same order of magnitude can be expected for the reduced range of  $\mu(R)$  for monatomic molecules. The  $\Omega$ -modulation effect therefore contributes only about 3% to the total binary absorption coefficient of the translational spectra of monatomic gas mixtures. An interpretation of the translational spectrum as a kind of continuous rotational spectrum due to the turning of the dipole moment during a collision is therefore not appropriate.

At high temperatures, the classical value of the pair distribution function,

$$(39) \quad g^{(0)}(x) = \exp(-V^*/T^*),$$

where  $V^* = V/\epsilon$  and  $T^* = kT/\epsilon$ , can be used. The corresponding classical values of the integral (38) can be obtained by numerical integration. For a number of reduced temperatures and values of  $\rho^* = \rho/\sigma$ , the classical values of the integral (38) are given in Table I. If experimental values of  $\alpha_1(T)$  are

TABLE I  
Classical values of the integrals  $I_1$ ,  $K$ , and  $J$

$T^*$	$\rho/\sigma$			$\rho/\sigma$			$J^{(0)} \times 10^{-2}$
	0.12	0.13	0.14	0.12	0.13	0.14	
	$I_1^{(0)} \times 10^{-2}$			$K^{(0)} \times 10^{-2}$			
1	1.54	1.42	1.32	4.27	4.24	4.21	3.44
2	1.49	1.32	1.19	3.74	3.65	3.57	2.74
3	1.70	1.47	1.29	3.99	3.84	3.72	2.72
4	1.95	1.65	1.43	4.35	4.15	3.99	2.81
5	2.20	1.83	1.57	4.74	4.49	4.29	2.92
6	2.46	2.02	1.71	5.14	4.83	4.59	3.05
7	2.71	2.21	1.85	5.53	5.17	4.89	3.18
8	2.97	2.39	1.99	5.92	5.51	5.18	3.31
9	3.22	2.57	2.12	6.30	5.84	5.47	3.43
10	3.47	2.75	2.26	6.67	6.16	5.76	3.56

available over a certain range of temperatures, empirical values of the parameters  $\rho^*$ ,  $\lambda^2\gamma$ , and  $\epsilon$  can be obtained. As in the case of the second virial coefficient this can best be done by plotting on a transparent sheet  $\ln \alpha_1$  vs.  $\ln T$  and on another sheet  $\ln I_1$  vs.  $\ln T^*$  for a number of different values of  $\rho^*$ , and trying to make the best fit by shifting the two sheets relative to each

other parallel to the co-ordinate axes. The curve  $\ln I_1$  vs.  $\ln T^*$  giving the best fit provides the value of  $\rho^*$ ; the shift in the co-ordinate direction yields  $\epsilon$ ; and the shift in the ordinate direction gives  $\lambda^2\gamma$ .

The measurements of the translational spectra of mixtures of the rare gases, which have been published so far (Kiss and Welsh 1959a) do not extend to sufficiently low frequencies to make a determination of  $\alpha_1$  from the experimental data possible. An extension of these measurements to lower frequencies is therefore required to obtain experimental values of the parameters  $\rho^*$ ,  $\epsilon$ , and  $\lambda^2\gamma$ . In this connection a theoretical calculation of the induced dipole moment in terms of the electronic structure of the molecules would be of great interest.

### 5. THE BINARY ABSORPTION COEFFICIENTS OF DIATOMIC GASES AND OF DIATOMIC-MONATOMIC GAS MIXTURES

We first consider a single component diatomic gas such as hydrogen. The binary absorption coefficient of the far infrared spectrum which now comprises a translational and a rotational band is given by (14). The trace appearing in this expression contains a trace over the motion of the center of mass of the pair of molecules, which can be carried out, giving

$$(40) \quad \alpha_1 = \frac{1}{2}(\kappa\lambda^3/Z^2)\text{Tr}\{\exp(-\beta H)\mathbf{y} \cdot [K, \mathbf{y}]\}.$$

$\lambda = h/(2\pi mkT)^{\frac{1}{2}}$ , where  $m$  is the reduced mass of the pair of molecules, and  $H$  and  $K$  denote the total and kinetic energies respectively of the relative translational motion and of the rotational motion of the molecules. The effect of the anisotropic intermolecular forces, which appear only in the Boltzmann factor and not in the commutator with  $\mathbf{y}$ , is small as long as these forces do not mix appreciably the states of different rotational angular momentum of the molecules. We neglect the anisotropic intermolecular potential in the Boltzmann factor in (40). The quantity  $Z$  is then equal to the rotational partition function

$$(41) \quad Z = \sum_J g_J \exp(-\beta E_J),$$

of a molecule. The anisotropic forces produce a coupling between the translational and rotational motions of the molecules, but this coupling does not affect the binary absorption coefficient because of the invariance of the trace appearing in (40) (cf. Van Kranendonk 1957).

The pair dipole moment  $\mathbf{y}$  appearing in (40) is a function of the orientations of the intermolecular axis  $\mathbf{R}$  and of the internuclear axes of the molecules 1 and 2. The orientations of these three axes can be specified relative to a co-ordinate system fixed in space, and these orientations are then denoted by  $\Omega \equiv (\theta, \phi)$ ,  $\Omega_1 \equiv (\theta_1, \phi_1)$ , and  $\Omega_2 \equiv (\theta_2, \phi_2)$ . The orientations of the internuclear axes can also be specified relative to the intermolecular axis  $\mathbf{R}$  as polar axis, and are then denoted by  $\omega_1$  and  $\omega_2$ . The components,  $\mu_\epsilon$ , of  $\mathbf{y}$  in a co-ordinate system fixed to  $\mathbf{R}$  can be expanded in terms of products of spherical harmonics of  $\omega_1$  and  $\omega_2$ ; the corresponding expansion coefficients are denoted by  $C_s(\lambda_1 \mu_1 \lambda_2 \mu_2; R)$  (Van Kranendonk and Kiss 1959). However, in

in the present context it is more convenient to work with the components,  $\mu_s$ , of  $\mathbf{y}$  in a co-ordinate system fixed in space, since the motion of the intermolecular axis  $\mathbf{R}$  is now also of importance. The dependence of  $\mu_s$  on  $\Omega$ ,  $\Omega_1$ , and  $\Omega_2$  can be expanded in terms of the complete set of eigenfunctions

$$(42) \quad \psi(J_1 J_2 L, J_3 J_m; \Omega, \Omega_1, \Omega_2),$$

of  $\mathbf{J}_1^2$ ,  $\mathbf{J}_2^2$ ,  $\mathbf{L}^2$ ,  $\mathbf{J}_3^2$ ,  $\mathbf{J}^2$ , and  $J_z$ , where  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are the rotational angular momenta, and  $\mathbf{L}$  is the angular momentum of the relative translational motion of the molecules;  $\mathbf{J}_3 = \mathbf{J}_1 + \mathbf{J}_2$  and  $\mathbf{J} = \mathbf{J}_3 + \mathbf{L}$ . The expansion of the spherical components  $\mu_s$ ,  $s = 0, \pm 1$ , defined as

$$(43) \quad \begin{cases} \mu_0 = \mu_z, \\ \mu_{\pm 1} = \mp 2^{-\frac{1}{2}}(\mu_x \pm i\mu_y), \end{cases}$$

takes the form

$$(44) \quad \mu_s = (64\pi^3/3)^{\frac{1}{2}} \sum_{\lambda_1 \lambda_2 \Lambda L} A(\lambda_1 \lambda_2 \Lambda L; R) \psi(\lambda_1 \lambda_2 L, \Lambda L; \Omega, \Omega_1, \Omega_2).$$

The expansion coefficients  $A(\lambda_1 \lambda_2 \Lambda L; R)$  are independent of the co-ordinate system, and provide an invariant classification of the induction effects in terms of the parameters  $\lambda_1$ ,  $\lambda_2$ ,  $\Lambda$ , and  $L$ . The values of these parameters are restricted by the triangular conditions (cf. Rose 1957),

$$(45) \quad \Delta(\lambda_1 \lambda_2 \Lambda) \quad \text{and} \quad \Delta(\Lambda L),$$

and by the fact that  $\lambda_1 + \lambda_2 + \Lambda + 1$  must be even. This last property follows from the transformation properties of  $\mu_s$  under a reflection in a plane through the origin. Finally, one can show that for the usual choice of phases of the functions (42), the coefficients  $A(\lambda_1 \lambda_2 \Lambda L; R)$  are all real. If the molecules 1 and 2 are identical, we have the additional symmetry property

$$(46) \quad A(\lambda_1 \lambda_2 \Lambda L; R) = (-1)^{\Lambda+1} A(\lambda_2 \lambda_1 \Lambda L; R).$$

The relation between the coefficients  $A(\lambda_1 \lambda_2 \Lambda L)$  and  $C_s(\lambda_1 \mu_1 \lambda_2 \mu_2)$  is given by

$$(47) \quad A(\lambda_1 \lambda_2 \Lambda L) = \left(\frac{2L+1}{3}\right)^{\frac{1}{2}} \sum_{\kappa \mu_1 \mu_2} C(\lambda_1 \lambda_2 \Lambda; \mu_1 \mu_2) C(\Lambda L; \kappa 0) C_s(\lambda_1 \mu_1 \lambda_2 \mu_2),$$

where  $C(\lambda_1 \lambda_2 \Lambda; \mu_1 \mu_2)$  and  $C(\Lambda L; \kappa 0)$  are Clebsch-Gordan coefficients (cf. Rose 1957).

In the far infrared spectrum arising from the dipole moment (44), we can distinguish between four modulation effects, viz. the  $R$ ,  $\Omega$ ,  $\Omega_1$ , and  $\Omega_2$  effects which result from modulations occurring in the variables  $R$ ,  $\Omega$ ,  $\Omega_1$ , and  $\Omega_2$  during a collision. This separation of the total intensity is not unique as far as the modulations produced by the rotational motion of the molecules are concerned. It is also possible, for example, to consider a separation into  $R$ ,  $\Omega$ ,  $\omega_1$ , and  $\omega_2$  modulation effects. The separation of the binary absorption

coefficient (40) into the four contributions corresponding to the  $R$ ,  $\Omega$ ,  $\Omega_1$ , and  $\Omega_2$  modulation effects can be obtained by splitting the kinetic energy in (40) into the corresponding four terms,

$$(48) \quad K = -(\hbar^2/2m)[\partial^2/\partial R^2 + (2/R)\partial/\partial R] + (\hbar^2/2mR^2)\mathbf{L}^2 + (\hbar^2/2I)\mathbf{J}_1^2 + (\hbar^2/2I)\mathbf{J}_2^2,$$

where  $I$  is the moment of inertia of the molecules. When the expression (48) is substituted into (40), the resulting terms can be evaluated in a straightforward way. The result can be written in the form

$$(49) \quad \alpha_1 = (2\pi^2/3c) \sum_{\lambda_1 \lambda_2 \Delta L} \int_0^\infty \left[ \frac{1}{m} |A(\lambda_1 \lambda_2 \Delta L; R)'|^2 + \left\{ \frac{\lambda_1(\lambda_1+1)}{I} + \frac{\lambda_2(\lambda_2+1)}{I} + \frac{L(L+1)}{mR^2} \right\} |A(\lambda_1 \lambda_2 \Delta L; R)|^2 \right] g(R) R^2 dR,$$

where  $g(R)$  is the pair distribution function for isotropic intermolecular forces, and the prime denotes differentiation with respect to  $R$ . The expression (49) is the general expression for the binary absorption coefficient of the far infrared spectrum of a diatomic gas in terms of the expansion coefficients  $A(\lambda_1 \lambda_2 \Delta L; R)$  characterizing the induced pair dipole moment.

In the absence of anisotropic intermolecular forces, the far infrared spectrum can be split into a translational and a rotational band. The translational band is defined as the intensity arising from those transitions in which the total rotational energy of the molecules does not change. The binary absorption coefficient of the translational band defined in this way receives contributions from all those transitions in pairs of molecules for which  $\Delta J_1 = -\Delta J_2$ , where  $\Delta J_i = J'_i - J_i$  is the change in the rotational quantum number of molecule  $i$ . The contribution of these transitions to the binary absorption coefficient (40) can be calculated, and the result can be written in the form

$$(50) \quad \alpha_{1, \text{tr}} = (2\pi^2/3mc) \sum_{\lambda_1 \lambda_2 \Delta L} \sum_{\Delta J} L_{\lambda_1}(\Delta J) L_{\lambda_2}(-\Delta J) \int_0^\infty \left[ |A(\lambda_1 \lambda_2 \Delta L; R)'|^2 + \frac{L(L+1)}{R^2} |A(\lambda_1 \lambda_2 \Delta L; R)|^2 \right] g(R) R^2 dR,$$

where

$$(51) \quad L_\lambda(\Delta J) = Z^{-1} \sum_j g_j \exp(-\beta E_j) C(J, \lambda, J + \Delta J; 0, 0)^2,$$

$C(J, \lambda, J'; 0, 0)$  is a Clebsch-Gordan coefficient (cf. Rose 1957), and  $Z$  is defined by (41). The binary absorption coefficient of the rotational band,  $\alpha_{1, \text{rot}}$ , can be obtained by subtracting (50) from (49). We note that the second and third term in (49), proportional to  $\lambda_1(\lambda_1+1)$  and  $\lambda_2(\lambda_2+1)$  respectively, contribute only to  $\alpha_{1, \text{rot}}$  and not to  $\alpha_{1, \text{tr}}$ , since these terms are absent in (50). However, the sum of these two rotational terms and the translational term (50) is not equal to the total intensity (49), since in general  $\sum_{\Delta J} L_{\lambda_1}(\Delta J) L_{\lambda_2}(-\Delta J) < 1$ . The typical translational terms in (48), viz. the first and

fourth term containing the reduced mass  $m$ , thus give a non-vanishing contribution to the intensity of the rotational lines. This is due to the fact that we work with  $\alpha$  rather than with  $\tilde{\alpha}$ . Because of the translational effect, the rotational lines are broadened asymmetrically and the mean frequency in a rotational line is larger than the rotational frequency, and  $\alpha_{1,\text{rot}}$  is therefore larger than the sum of the pure rotational terms in (49). This concludes the derivation of the general expressions for the binary absorption coefficients of a diatomic gas.

We next consider a diatomic-monatomic gas mixture. The binary absorption coefficient  $\alpha'_1$  defined by (23) can be obtained from (49) by putting  $\lambda_2 = 0$  and multiplying by a factor 2. In this way we obtain

$$(52) \quad \alpha'_1 = (4\pi^2/3c) \sum_{\lambda L} \int_0^\infty \left[ \frac{1}{m} |A(\lambda 0 \lambda L; R)'|^2 + \left\{ \frac{\lambda(\lambda+1)}{I} + \frac{L(L+1)}{mR^2} \right\} |A(\lambda 0 \lambda L; R)'|^2 \right] g(R) R^2 dR,$$

where  $I$  is the moment of inertia of the diatomic molecule, and  $m$  is the reduced mass of a diatomic-monatomic pair. The corresponding expression for the translational band is given by

$$(53) \quad \alpha'_{1,\text{tr}} = (4\pi^2/3mc) \sum_{\lambda L} L_\lambda(0) \int_0^\infty \left[ |A(\lambda 0 \lambda L; R)'|^2 + \frac{L(L+1)}{R^2} |A(\lambda 0 \lambda L; R)'|^2 \right] g(R) R^2 dR.$$

The binary absorption coefficients (49), (50), (52), and (53) are functions of the temperature, which can be evaluated explicitly by introducing a model for the induced pair dipole moment, i.e. for the magnitude and  $R$ -dependence of the expansion coefficients  $A(\lambda_1 \lambda_2 \Delta L; R)$ . This will be discussed in the next section for the (exp-4) model which has also been used for the vibrational (Van Kranendonk 1958) and for the rotational (Van Kranendonk and Kiss 1959) spectra.

#### 6. EVALUATION OF THE BINARY ABSORPTION COEFFICIENTS OF DIATOMIC GASES AND OF DIATOMIC-MONATOMIC GAS MIXTURES IN TERMS OF THE (exp-4) MODEL

We first consider the translational spectrum of a pure diatomic gas. In contradistinction to the case of monatomic molecules, the induced dipole moment for diatomic molecules contains, in addition to the short-range part due to the overlap induction effect, a long-range part. We take into account only the part of longest range, which arises from the quadrupolar induction effect (Van Kranendonk 1958), and which varies as  $R^{-4}$ . We assume that the overlap part falls off exponentially with increasing separation  $R$ , and we take into account only the non-vanishing terms corresponding to the simplest angle-dependent overlap terms. This leads to the following set of non-vanishing coefficients  $A(\lambda_1 \lambda_2 \Delta L; R)$ ,

$$(54) \quad \begin{cases} A(2021) = -A(0221) = \xi_1 \exp(-R/\rho), \\ A(2023) = -A(0223) = \xi_3 \exp(-R/\rho) + 3^{\frac{1}{2}}(\alpha Q/R^4), \end{cases}$$

where  $Q$  is the quadrupole moment and  $\alpha$  the average polarizability of the molecules, and  $\xi_1$  and  $\xi_3$  are the parameters describing the overlap moment. We note that the coefficient  $A(0001)$  corresponding to the angle-independent overlap moment, which gives the main contribution in the case of two different molecules, is equal to zero here because of the symmetry relation (46). Substituting (54) into the expression (50) for the translational binary absorption coefficient,  $\alpha_{1,\text{tr}}$ , we find

$$(55) \quad \alpha_{1,\text{tr}} = [\lambda_1^2 I_1 + \lambda_3^2 I_3 + \mu^2 J + \mu \lambda_3 K] L_2(0) \gamma.$$

The dimensionless quantities  $\lambda_1$ ,  $\lambda_3$ , and  $\mu$  are defined as (cf. Van Kranendonk and Kiss 1959)

$$(56) \quad \lambda_{1,3} = (\xi_{1,3}/e\sigma) \exp(-\sigma/\rho), \quad \mu = (\alpha Q/e\sigma^5),$$

and the quantity

$$(57) \quad \gamma = (\pi e^2 \sigma^3 / 3mc)$$

has the dimension of a binary absorption coefficient. The integrals  $I_L$ ,  $J$ , and  $K$  are defined as

$$(58) \quad \begin{cases} I_L = 4\pi \int_0^\infty \exp\left[-\frac{2\sigma}{\rho}(x-1)\right] \left[ \left(\frac{\sigma}{\rho}\right)^2 + \frac{L(L+1)}{x^2} \right] g(x) x^2 dx, \\ J = 336\pi \int_0^\infty x^{-8} g(x) dx, \\ K = 32\pi\sqrt{3} \int_0^\infty \exp\left[-\frac{\sigma}{\rho}(x-1)\right] \left[ \frac{\sigma}{\rho} + \frac{3}{x} \right] g(x) x^{-3} dx, \end{cases}$$

where  $x = R/\sigma$ . The quantity  $L_2(0)$  is defined by (51). At high temperatures the classical value (39) of the pair distribution function may be used. At intermediate temperatures  $g(x)$  can be expanded in powers of Planck's constant (cf. De Boer 1948). Corresponding to this expression of  $g(x)$ , the integrals (58) can be written in the form

$$(59) \quad I_L = I_L^{(0)} + \Lambda^{*2} I_L^{(2)} + \Lambda^{*4} I_L^{(4)} + \dots,$$

and similarly for  $J$  and  $K$ , where  $\Lambda^* = (h^2/m\epsilon\sigma^2)^{\frac{1}{2}}$ . For a number of reduced temperatures and values of  $\rho/\sigma$  the classical values  $I_L^{(0)}$ ,  $J^{(0)}$ , and  $K^{(0)}$  of the integrals (58) are given in Table I. We note here, that the quantities  $\xi_1$  and  $\lambda_1$ , and the integrals  $I_L$ ,  $J$ , and  $K$  defined in this paper are not the same as those used in a previous publication (Van Kranendonk and Kiss 1959).

We now discuss the binary absorption coefficient of the translational spectrum of hydrogen at 300° K. We use the values  $\rho/\sigma = 0.126$ ,  $\lambda_3 = 0.6 \times 10^{-4}$ , and  $\lambda_1 = 10^{-4}$  obtained from the rotational spectrum (Van Kranendonk and Kiss 1959). For the expectation value of the quadrupole moment over the ground vibrational state we use the theoretical value  $Q = (0.46 + 0.03)ea_0^2$

obtained from the calculations of Kolos and Roothaan (1960), the second term representing the correction due to the anharmonicity of the vibrational motion. For the average polarizability the experimental value  $\alpha = 5.44a_0^3$  (Essen 1953) is used. It turns out that the contributions from the terms  $\lambda_1^2$  and  $\lambda_2^2$  in (55) are of the order of 1% of the term  $\mu^2$ , and we therefore neglect these terms. For the binary absorption coefficient of equilibrium hydrogen at 300° K we find, after inserting the numerical values mentioned above,

$$(60) \quad \alpha'_{1,\text{tr}} = (3.3 + 0.7) \times 10^{-32} \text{ sec}^{-1} \text{ cm}^5.$$

The first term in (60) is the intensity due to the quadrupolar induction effect, and corresponds to the term  $\mu^2$  in (55). The second term in (60) represents the interference effect between the quadrupolar and overlap moments, and corresponds to the term  $\mu\lambda_3$  in (55). Because of the uncertainty in the value of  $\lambda_3$  derived from the rotational spectrum, it would be interesting if the value (60) could be checked experimentally by extending the existing measurements of the translational spectrum of hydrogen (Kiss and Welsh 1959b) to lower frequencies.

For diatomic-monatomic mixtures the quadrupolar induction effect is characterized by the coefficient

$$(61) \quad A(2023) = 3^{\frac{1}{2}} Q_1 \alpha_2 R^{-4},$$

where  $Q_1$  is the quadrupole moment of the diatomic molecule, and  $\alpha_2$  is the polarizability of the monatomic molecule. For two different molecules, the relation (46) no longer holds, and an angle-independent overlap moment of the type  $A(0001)$  appears. As this term gives the main contribution to the overlap moment we neglect the angle-dependent part, and we put

$$(62) \quad A(0001) = \xi \exp(-R/\rho).$$

Substituting the expansion coefficients (61) and (62) into the expression (53) for  $\alpha'_{1,\text{tr}}$ , we find

$$(63) \quad \alpha'_{1,\text{tr}} = [\lambda^2 I_1 + \mu^2 J L_2(0)] \gamma,$$

where

$$(64) \quad \lambda = (\xi/e\sigma) \exp(-\sigma/\rho), \quad \mu = (Q_1 \alpha_2/e\sigma^5),$$

and the remaining quantities are defined as in (51), (57), and (58). From experiments on the translational absorption of diatomic-monatomic gas mixtures as a function of temperature the parameters  $\lambda$  and  $\rho/\sigma$  can be found.

#### 7. CONCLUDING REMARKS

The theory of the translational absorption bands in gases developed in this paper should be compared to the theory of the equation of state of gases. The binary absorption coefficients introduced here take the place of the second virial coefficients in the equation of state, and the quantity which can be determined from the experimental data is the induced dipole moment rather than the intermolecular potential. The expressions for the binary

absorption coefficients have the same general validity as those for the second virial coefficients, the main assumption being that the electrons follow the motion of the nuclei adiabatically. The ultimate goal of the theory is to calculate the induced dipole moments in terms of the electronic structure of the molecules, and to obtain in this way in combination with the experimental results new information about the intermolecular interaction.

#### ACKNOWLEDGMENTS

The authors wish to express their gratitude to Professor W. H. Watson for placing the facilities of the Computation Center of the University of Toronto at their disposal.

#### REFERENCES

CHISHOLM, D. A. and WELSH, H. L. 1954. *Can. J. Phys.* **32**, 291.  
CRAWFORD, M. F., WELSH, H. L., MACDONALD, J. C. F., and LOCKE, J. L. 1950. *Phys. Rev.* **80**, 469.  
ESSEN, L. 1953. *Proc. Phys. Soc. B*, **66**, 189.  
GUSH, H. P., HARE, W. F. J., ALLIN, E. J., and WELSH, H. L. 1960. *Can. J. Phys.* **38**, 176.  
HIRCHFELDER, J. O., CURTISS, C. F., and BIRD, R. B. 1954. *Molecular theory of gases and liquids* (John Wiley & Sons, Inc., New York).  
KISS, Z. J. and WELSH, H. L. 1959a. *Phys. Rev. Letters*, **2**, 166.  
——— 1959b. *Can. J. Phys.* **37**, 1249.  
KOLOS, W. and ROOHAAN, C. C. J. 1960. *Rev. Modern Phys.* **32**, 219.  
ROSE, M. E. 1957. *Elementary theory of angular momentum* (John Wiley & Sons, Inc., New York).  
UHLENBECK, G. E. and BETH, E. 1936. *Physica*, **3**, 729.  
VAN KRANENDONK, J. 1957. *Physica*, **23**, 825.  
——— 1958. *Physica*, **24**, 347.  
VAN KRANENDONK, J. and KISS, Z. J. 1959. *Can. J. Phys.* **37**, 1187.

## PREDISSOCIATION IN THE HNO MOLECULE<sup>1</sup>

M. J. Y. CLEMENT<sup>2</sup> and D. A. RAMSAY

### ABSTRACT

Twelve bands of HNO and 18 bands of DNO in the region 6,000 to 10,000 Å have been photographed in emission during the reactions of hydrogen and deuterium atoms with nitric oxide. Two of the HNO bands and 3 of the DNO bands show a sharp breaking-off in the *K*-rotational structure, due to predissociation of the molecule in the excited state. Upper limits for the dissociation energies of HNO and DNO are 48.6 kcal/mole and 49.1 kcal/mole respectively.

Cashion and Polanyi (1959) have shown that bands of HNO can be detected in emission during the reaction of hydrogen atoms with nitric oxide. We have recently studied the reactions H+NO and D+NO and have observed several emission bands of HNO and DNO in the region 6,000–10,000 Å using (i) a three-prism glass spectrograph with *f*/2 and *f*/10 cameras and (ii) an *f*/4 grating spectrograph with a dispersion of 5 Å/mm. The most interesting new result which has emerged from the present studies is that some of the HNO and DNO bands show a distinct breaking-off in the *K*-rotational structure, due to predissociation in the excited state. These results provide the first example of the breaking-off in the rotational structure in the spectrum of a polyatomic molecule and afford an upper limit for the energy of the reaction H+NO → HNO.

The apparatus consisted of a reaction tube 5 cm in diameter and 30 cm in length. Hydrogen atoms were pumped from Wood's tubes through three short side-arms 2.5 cm in diameter and situated 7.5 cm apart. Nitric oxide was introduced through three smaller side-arms situated diametrically opposite. The radiation emitted by reaction H+NO was not visible but was sufficiently intense that the (000)–(000) band of HNO near 7600 Å could be photographed with the three-prism spectrograph and *f*/2 camera in 10 seconds using unsensitized Kodak I-N plates. The intensity of the emission was a maximum with partial pressures of the reagents  $\sim$ 1 mm Hg. Final plates of HNO and DNO were taken with the *f*/10 camera using hypersensitized I-N and I-M plates. Reproductions of the spectra in the region 6000 to 9000 Å are shown in Fig. 1. Some of the stronger bands were also photographed with the *f*/4 grating spectrograph but a discussion of the rotational analyses of these bands will be deferred to a later paper.

The wavelengths and probable assignments of the HNO and DNO bands are given in Tables I and II. The values for  $\nu_{\text{calc}}$  were obtained using the following data:

<sup>1</sup>Manuscript received October 11, 1960.

Contribution from the Division of Pure Physics, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 6105.

<sup>2</sup>N.R.C. Summer Research Student 1960. Present address: Department of Physics, University of British Columbia, Vancouver, B.C.

HNO	$T'_{00} = 13\ 154.38\text{ cm}^{-1}$	$\nu'_2 = 1420.77\text{ cm}^{-1}$	$\nu'_3 = 981.18\text{ cm}^{-1}$
	$\nu'_1 = 3\ 596\text{ cm}^{-1}$	$\nu'_{2'} = 1562\text{ cm}^{-1}$	$\nu'_{3'} = 1110\text{ cm}^{-1}$
DNO	$T'_{00} = 13\ 180.32\text{ cm}^{-1}$	$\nu'_2 = 1401.28\text{ cm}^{-1}$	$\nu'_3 = 755.31\text{ cm}^{-1}$
	$\nu'_1 = 2\ 680\text{ cm}^{-1}$	$\nu'_{2'} = 1550\text{ cm}^{-1}$	$\nu'_{3'} = 820\text{ cm}^{-1}$

TABLE I  
Emission features of HNO

$\lambda_{\text{air}}(\text{\AA})$	Intensity	$\nu_{\text{vac}}(\text{cm}^{-1})$	Assignment	$\nu_{\text{obs}}(\text{cm}^{-1})$		
				Cashion and Polanyi	Dalby	$\nu_{\text{calc}}(\text{cm}^{-1})$
9580 $\pm$ 50	w	10 435 $\pm$ 55	(000)-(011)	10 530 $\pm$ 30		10 482 <sup>c</sup>
9275 $\pm$ 30	vw	10 780 $\pm$ 35	(020)-(110)?			10 838 <sup>c</sup>
			(010)-(012)?			10 793 <sup>c</sup>
9120 $\pm$ 40	m	10 960 $\pm$ 50	(010)-(100)			10 979 <sup>c</sup>
			(000)-(002)?			10 934 <sup>c</sup>
8850 $\pm$ 40	w	11 295 $\pm$ 50	(020)-(030)?			11 310 <sup>c</sup>
8650 $\pm$ 80 <sup>a</sup>	m	11 560 $\pm$ 105	(000)-(010)	11 370 $\pm$ 60		11 592
8265 $\pm$ 40 <sup>a</sup>	w	12 095 $\pm$ 60	(000)-(001)	12 060 $\pm$ 30		12 044
7965 $\pm$ 70 <sup>a</sup>	m	12 550 $\pm$ 110	(100)-(100)			
7625 $\pm$ 100 <sup>a</sup>	vs	13 110 $\pm$ 170	(000)-(000)	13 050 $\pm$ 60	13 154	
7379.2		13 547.9	R-head 7-6			13 547.7
7341.3		13 617.8	R-head 8-7			13 617.2
7303.2		13 688.9	R-head 9-8			
7266.3		13 758.4	R-head 10-9			
7230.1		13 827.2	R-head 11-10			
7195.5		13 893.7	R-head 12-11			
7162.2		13 958.3	R-head 13-12			
7100 $\pm$ 30	w	14 080 $\pm$ 60	(001)-(000)		14 135	
6925 $\pm$ 60	m	14 435 $\pm$ 125	(020)-(010)?	14 420 $\pm$ 50		14 434 <sup>c</sup>
			(010)-(000)?			
6797.4		14 707.4	R-head 3-2			14 575
6771.5		14 763.7	R-head 4-3			14 707.4 <sup>b</sup>
6744.3		14 823.2	R-head 5-4			14 764.4 <sup>b</sup>
6715.7		14 886.4	R-head 6-5			14 824.0 <sup>b</sup>
6685.7		14 953.2	R-head 7-6			14 886.8 <sup>b</sup>
6656.2		15 019.4	R-head 8-7			
6627.8		15 083.8	R-head 9-8			
6598.4		15 151.0	R-head 10-9			
6453 $\pm$ 25	vw	15 492 $\pm$ 60	(011)-(000)			15 556 <sup>c</sup>
6272 $\pm$ 50	vw	15 940 $\pm$ 125	(020)-(000)			15 996 <sup>c</sup>

<sup>a</sup>Observed also with f/4 grating spectrograph (dispersion 5 Å/mm).

<sup>b</sup>These lines are the last lines observed by Dalby and lie close to the band heads.

<sup>c</sup>Neglecting anharmonicity.

The constants for the excited state are those derived by Dalby (1958) while those for the ground state are taken from the measurements of Harvey and Brown (1959) on the infrared spectrum of solid HNO and DNO (see also Brown and Pimentel 1958). The agreement between the observed and calculated frequencies is seen to be satisfactory and suggests that most of the assignments are probably correct.

Our observations of the HNO bands agree fairly well with those of Cashion and Polanyi except that we have not so far been able to confirm the band which they reported at 16 200 cm<sup>-1</sup> (6171 Å); moreover we observed a band of medium intensity at 7965 Å which was not reported by these authors. The latter band has also been photographed under higher dispersion (5 Å/mm)

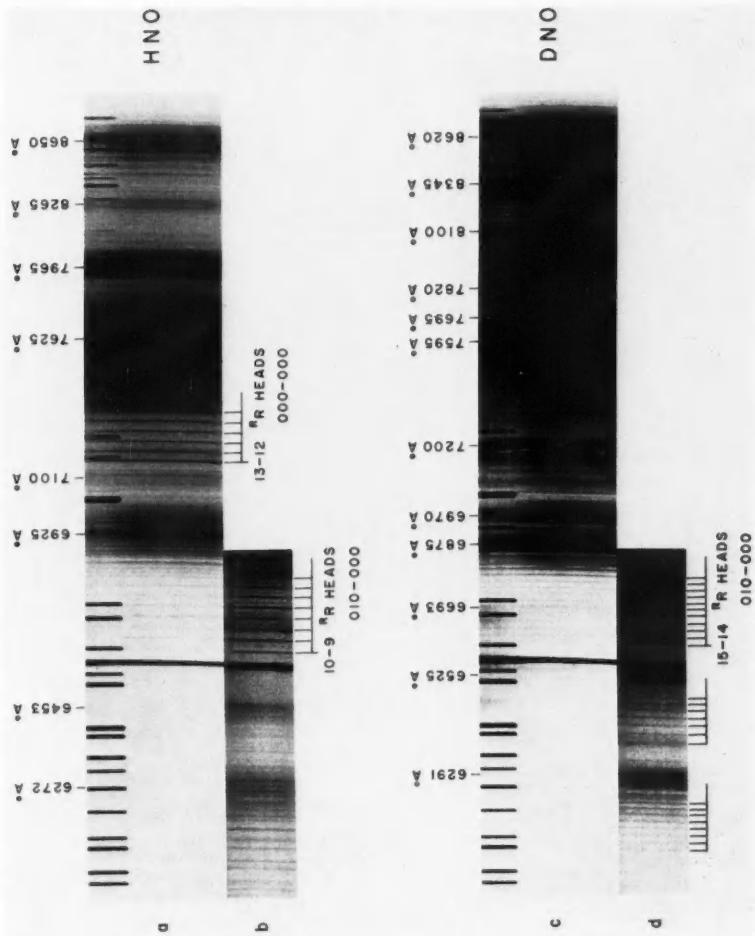


FIG. 1. Emission bands of HNO and DNO photographed with a three-prism glass spectrograph,  $f/10$  camera, and hypersensitized I-N plates. The exposure times are (a) 20 min, (b) 2 hr, (c) 20 min, (d) 3 hr. A neon lamp was used to provide the reference spectrum.

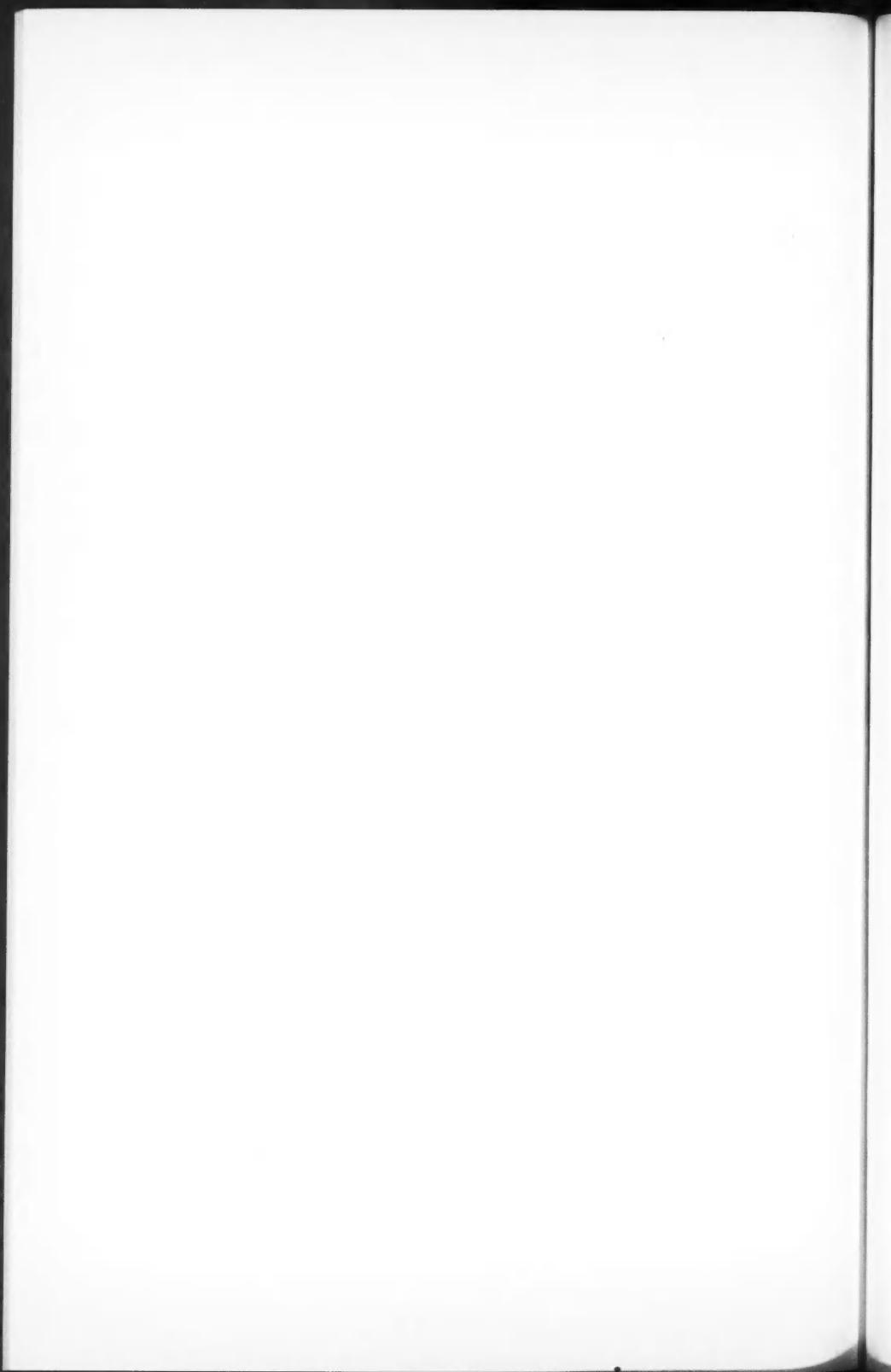


TABLE II  
Emission features of DNO

$\lambda_{\text{air}}(\text{\AA})$	Intensity	$\nu_{\text{vac}}(\text{cm}^{-1})$	Assignment	$\nu_{\text{obs}}(\text{cm}^{-1})^a$	$\nu_{\text{calc}}(\text{cm}^{-1})$
9570 $\pm$ 35	vw	10 445 $\pm$ 40	(000)–(100)		10 500
9255 $\pm$ 60	w	10 800 $\pm$ 70	(000)–(011)		10 810 <sup>d</sup>
8990 $\pm$ 30	vw	11 120 $\pm$ 35	?		
8915 $\pm$ 30	vw	11 215 $\pm$ 40	?		
8840 $\pm$ 30	w	11 310 $\pm$ 40	(020)–(030)		11 333 <sup>d</sup>
8715 $\pm$ 30	w	11 470 $\pm$ 40	(010)–(020)		11 482 <sup>d</sup>
8620 $\pm$ 30 <sup>b</sup>	m	11 600 $\pm$ 40	(000)–(010)		11 630
8345 $\pm$ 30 <sup>b</sup>	m	11 980 $\pm$ 45	(010)–(100)		11 902
8100 $\pm$ 20	w	12 340 $\pm$ 30	(000)–(001)		12 360
7820 $\pm$ 50 <sup>b</sup>	m	12 785 $\pm$ 80	(100)–(100)		
7695 $\pm$ 30 <sup>b</sup>	m	12 990 $\pm$ 50	(010)–(010)		13 032
7595 $\pm$ 60 <sup>b</sup>	s	13 165 $\pm$ 105	(000)–(000)	13 180	
7200 $\pm$ 25 <sup>b</sup>	w	13 885 $\pm$ 50	(001)–(000)	13 936	
6970 $\pm$ 20 <sup>b</sup>	w	14 345 $\pm$ 40	(020)–(010)		14 433 <sup>d</sup>
6875 $\pm$ 20 <sup>b</sup>	w	14 540 $\pm$ 40	(010)–(000)	14 582	
6820.7		14 657.2	R-head 3–2	14 658.4 <sup>e</sup>	
6806.2		14 688.4	R-head 4–3	14 691.8 <sup>e</sup>	
6789.9		14 723.7	R-head 5–4	14 726.0 <sup>e</sup>	
6773.9		14 758.5	R-head 6–5	14 762.6 <sup>e</sup>	
6755.3		14 799.1	R-head 7–6	14 801.1 <sup>e</sup>	
6738.2		14 836.7	R-head 8–7	14 841.1 <sup>e</sup>	
6717.6		14 882.2	R-head 9–8	14 882.2	
6699.9		14 921.5	R-head 10–9		
6679.7		14 966.6	R-head 11–10		
6663.0		15 004.1	R-head 12–11		
6644.5		15 045.9	R-head 13–12		
6627.3		15 084.9	R-head 14–13		
6608.9		15 126.9	R-head 15–14		
6693 $\pm$ 15	vw	14 937 $\pm$ 35			
6525 $\pm$ 20	vw	15 321 $\pm$ 45	(011)–(000)		15 337 <sup>d</sup>
6476.0		15 437.4	R-head		
6461.7		15 471.5	R-head		
6446.0		15 509.2	R-head		
6428.8		15 550.7	R-head		
6410.1		15 596.1	R-head		
6391.2		15 642.2	R-head		
6370.5		15 693.0	R-head		
6291 $\pm$ 20	vw	15 891 $\pm$ 50	(020)–(000)		15 983 <sup>d</sup>
6241.5		16 017.4	R-head		
6229.6		16 048.0	R-head		
6217.2		16 080.0	R-head		
6203.1		16 116.5	R-head		
6190.8		16 148.5	R-head		
6175.8		16 187.8	R-head		
6160.3		16 228.5	R-head		
6146.9		16 263.9	R-head		

<sup>a</sup>From the absorption measurements of Dalby (1958).<sup>b</sup>Observed also with f/4 grating spectrograph (dispersion 5 Å/mm).<sup>c</sup>These lines are the last lines observed by Dalby and lie close to the band heads.<sup>d</sup>Neglecting anharmonicity.

and has a rotational fine structure very similar to that of the (000)–(000) band of HNO. The assignment of this band is of particular interest since it would appear that the only reasonable assignment of the band is (100)  $\rightarrow$  (100). If this is correct, we may deduce that  $\nu'_1 - \nu''_1 \sim -560 \text{ cm}^{-1}$ . For DNO the corresponding band is at 7820 Å, and  $\nu'_1 - \nu''_1 \sim -380 \text{ cm}^{-1}$ .

The K-rotational structure of the bands is clearly visible on the spectra taken with the prism spectrograph and f/10 camera and the R-heads of the

subbands with  $\Delta K = +1$  appear as distinct linelike features. The measurements of the band heads agree with those of Dalby to within  $\sim 1 \text{ \AA}$ , in regions of overlap, and permit an unambiguous assignment of the rotational quantum numbers. For the (000)-(000) band of HNO a series of band heads is observed with good intensity up to the *R*-head of the  $K' = 13 \rightarrow K'' = 12$  subband. The series then terminates abruptly and the *R*-head of the  $K' = 14 \rightarrow K'' = 13$  subband is absent (see Fig. 1a). For the (010)-(000) band of HNO the corresponding series of band heads is observed up to  $K' = 10 \rightarrow K'' = 9$ . The *R*-head of the  $K' = 11 \rightarrow K'' = 10$  subband may be obscured by the  $H\alpha$  line but the *R*-head of the  $K' = 12 \rightarrow K'' = 11$  subband is definitely absent (see Fig. 1b). For the (010)-(000) band of DNO the last *R*-head observed before the series terminates is the *R*-head of the subband  $K' = 15 \rightarrow K'' = 14$  (see Fig. 1d). The total energies (electronic + vibrational + rotational) of the various levels in the excited state relative to ground states of HNO and DNO were calculated using the molecular constants given by Dalby and are as follows:

HNO: (000) level $K' = 13$ (observed)	$E = 16\ 440 \text{ cm}^{-1}$
$K' = 14$ (absent)	$E = 17\ 000 \text{ cm}^{-1}$
(010) level $K' = 10$ (observed)	$E = 16\ 500 \text{ cm}^{-1}$
$K' = 11$ (uncertain)	$E = 16\ 890 \text{ cm}^{-1}$
$K' = 12$ (absent)	$E = 17\ 320 \text{ cm}^{-1}$
DNO: (010) level $K' = 15$ (observed)	$E = 16\ 890 \text{ cm}^{-1}$
$K' = 16$ (absent)	$E = 17\ 190 \text{ cm}^{-1}$

The consistency between the three sets of values leaves no doubt that the breaking-off in the rotational structure is a genuine phenomenon. The breaking-off energy for DNO is expected to be slightly higher than for HNO, due to the difference in zero-point energies ( $\sim 600 \text{ cm}^{-1}$ ) for the two molecules. The breaking-off in the rotational structure can also be seen in the  $6525 \text{ \AA}$  and  $6291 \text{ \AA}$  bands of DNO (see Fig. 1d) but no evaluation of the appropriate energies can at present be made since no rotational analyses for these bands have yet been carried out.

The phenomenon of breaking-off in the rotational structure of bands is generally due to dissociation or predissociation of the molecule in the excited state (see Herzberg 1950). It is extremely unlikely that for HNO the phenomenon is due to dissociation since the breaking-off energy lies only  $3600 \pm 250 \text{ cm}^{-1}$  above the (000) level of the excited state. It is much more likely that the phenomenon is due to predissociation and that the potential curves representing the energy of the system as a function of the H-NO distance, are of the type shown in Fig. 2. A similar potential energy diagram has already been applied to the HCO molecule (Herzberg and Ramsay 1955) and in all probability a similar diagram applies for  $\text{HO}_2$ .

From Fig. 2 and the data given above, it is seen that for the dissociation of HNO and DNO into ground state products,

$$D''_0 (\text{HNO}) < 17\ 000 \text{ cm}^{-1} (48.6 \text{ kcal/mole, 2.11 ev}),$$

$$D''_0 (\text{DNO}) < 17\ 190 \text{ cm}^{-1} (49.1 \text{ kcal/mole, 2.13 ev}).$$

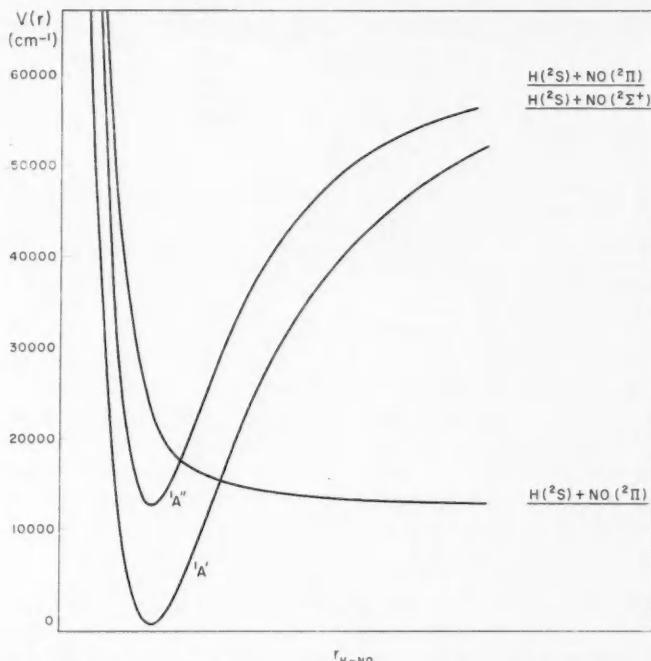


FIG. 2. Potential energy diagram for HNO. The combination  $H(^2S) + NO(^2\Pi)$  yields four states viz.  $^1A'$ ,  $^1A''$ ,  $^3A'$ ,  $^3A''$ , all of which are probably repulsive. There will be avoided crossings not indicated in the diagram, between the repulsive  $^1A'$  and  $^1A''$  states and the stable  $^1A'$  and  $^1A''$  states respectively.

#### REFERENCES

BROWN, H. W. and PIMENTEL, G. C. 1958. *J. Chem. Phys.* **29**, 883.  
 CASHION, J. K. and POLANYI, J. C. 1959. *J. Chem. Phys.* **30**, 317.  
 DALBY, F. W. 1958. *Can. J. Phys.* **36**, 1336.  
 HARVEY, K. B. and BROWN, H. W. 1959. *J. chim. phys.* **56**, 745.  
 HERZBERG, G. 1950. *Spectra of diatomic molecules* (D. Van Nostrand Co. Inc., Princeton, N.J.).  
 HERZBERG, G. and RAMSAY, D. A. 1955. *Proc. Roy. Soc. A*, **233**, 34.

#### NOTE ADDED IN PROOF

A slight, but definite, diffuseness of some of the rotational lines of the (021)-(000) band of HNO near 5900 Å in absorption has been found in recent flash photolysis experiments by Hollas and Ramsay (to be published). These new results confirm the predissociation limit established above.

# THE ABSORPTION SPECTRUM AND DISSOCIATION ENERGY OF SH<sup>1</sup>

J. W. C. JOHNS<sup>2</sup> AND D. A. RAMSAY

## ABSTRACT

The (2,0) bands of the  $A^2\Sigma^+ - X^2\Pi$  system of SH and SD have been photographed for the first time. More accurate values for the vibrational constants of the  $A^2\Sigma^+$  state have been obtained. The dissociation energy of SH in the excited state is  $D'_0 = 8020 \pm 1000 \text{ cm}^{-1}$  from which it is possible to deduce that the ground state dissociation energy  $D''_0$  (SH) is  $28,480 \pm 1000 \text{ cm}^{-1}$  ( $81.4 \pm 2.9 \text{ kcal/mole}$ ,  $3.53 \pm 0.12 \text{ ev}$ ).

## INTRODUCTION

It is well known from earlier work (Porter 1950; Ramsay 1952) that SH has a fairly large dissociation energy ( $D''_0 \sim 29,000 \text{ cm}^{-1}$ ) in its  $^2\Pi$  ground state, but a much smaller dissociation energy ( $D'_0 \sim 8000 \text{ cm}^{-1}$ ) in its first excited  $^2\Sigma^+$  state. Since the dissociation products of the  $^2\Sigma^+$  state are known with reasonable certainty [ $\text{SH}(^2\Sigma^+) \rightarrow \text{H}(^2\text{S}) + \text{S}(^1\text{D})$ ], the dissociation energy for the ground state may be derived from the dissociation energy of the excited state using the relation

$$(1) \quad D''_0(^2\Pi) = D'_0(^2\Sigma^+) + \nu_{00}(^2\Sigma^+ - ^2\Pi) - \nu_s(^1D - ^3P).$$

The (0,0) and (1,0) bands of the  $^2\Sigma^+ - ^2\Pi$  system of SH and SD have been observed in absorption by Porter and by Ramsay and the vibrational constants  $\omega_e'$  and  $\omega_e'x_e'$  for the excited state deduced with the aid of isotope relations. A value for  $D'_0$  was then obtained from a modified Birge-Sponer extrapolation. In the present work we have observed the (2,0) bands of SH and SD in absorption and have been able to determine the vibrational constants for the excited state without invoking isotope relations. Using a shorter and more reliable extrapolation, we have been able to derive a more accurate value for  $D'_0$  and hence, from equation (1), we obtain a more reliable value for the dissociation energy of the molecule in its ground state.

## EXPERIMENTAL

The new bands of SH and SD were observed during the flash photolysis of  $\text{H}_2\text{S}$  and  $\text{D}_2\text{S}$  at pressures of  $\sim 200 \text{ mm Hg}$ . The flash photolysis apparatus was similar to, though somewhat larger than, the apparatus described in earlier publications (for example, Callomon and Ramsay 1957). The quartz reaction tube was 2 m long and was fitted with multiple reflection mirrors of the type described by White (1942). Only the central half of the reaction tube was irradiated during the photolysis flash since sulphur is produced in the

<sup>1</sup>Manuscript received October 11, 1960.

Contribution from the Division of Pure Physics, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 6102.

<sup>2</sup>N.R.C. Postdoctorate Fellow 1959-60.

TABLE I  
Vacuum wavenumbers and assignments of the rotational lines of the (1,0) SH band

<i>J</i>	<i>P</i> <sub>1</sub>	<i>Q</i> <sub>1</sub>	<i>R</i> <sub>1</sub>	<i>qP</i> <sub>21</sub>	<i>qQ</i> <sub>21</sub>	<i>sR</i> <sub>21</sub>	<i>P</i> <sub>2</sub>	<i>Q</i> <sub>2</sub>	<i>R</i> <sub>2</sub>	<i>oP</i> <sub>12</sub>	<i>PQ</i> <sub>12</sub>	<i>qR</i> <sub>12</sub>
0.5	32,617.1*	32,632.6†	32,664.2††	32,632.6†	32,663.5††	32,709.71	32,234.9*	32,264.6*	32,294.86	32,220.1*	32,249.0*	†
1.5	586.81	618.1*	664.2††	617.1*	663.5††	725.41	218.0*	264.6*	325.46	187.63	218.0*	†
2.5	553.7*	600.3*	662.8††	599.3*	660.8††	738.06	196.8*	258.80	334.85	151.31	196.8*	32,260.02
3.5	517.30	579.3*	656.80	578.4*	655.33	747.59	171.1*	249.0*	340.27	111.0*	171.9*	250.82
4.5	478.17	555.48	648.45	553.7*	646.55	753.87	141.79	235.9*	367.38	143.26	237.28	
5.5	436.08	528.74	636.85	526.93	634.3*	756.9*	109.00	218.0*	019.66	111.0*	220.1*	
6.5	390.95	498.80	622.07	496.66	756.9*	072.29	196.8*	31,968.48	074.54	199.32		
7.5	342.71	465.71	603.94	463.14	756.9*	031.71	171.9*	034.25				
8.5	291.66	429.36	426.43									

\*Overlapped.

†Badly overlapped.

‡Taken from earlier measurements.

reaction, and the "dead space" at each end of the tube helped to prevent deposition of the sulphur on the mirrors. The mirrors were coated with alumnum and four traversals of the tube were used.

The photolysis lamp was 1 m long and was constructed in the manner recommended by Claesson and Lindquist (1957). The lamp was filled with argon at 1-2 cm pressure and was fired by a 120- $\mu$  condenser bank charged to 9 kv. The half-life of the flash was  $\sim 30$   $\mu$ sec. Absorption spectra were photographed using a quartz capillary flash lamp (3  $\mu$ f, 15 kv) as a source of continuum. The time delay between the photolysis and "source" flashes was kept as short as possible ( $\sim 5$   $\mu$ sec peak to peak) so that the SH bands could be photographed before absorption by  $S_2$  became troublesome. The spectra were photographed using a 21-ft concave grating spectrograph in the third and fourth orders and Eastman Kodak I-O plates. Approximately 40 flashes were needed for a satisfactory exposure. Reproductions of the (1,0) and (2,0) bands of SH and SD are given in Figs. 1 and 2. The rotational lines are distinctly diffuse due to predissociation of the molecule in the excited state.

An iron hollow cathode lamp was used to provide a reference spectrum. Standard wavelengths were taken from the M.I.T. wavelength tables (Harrison 1939) and were combined with the vacuum corrections of Edlén (1953). The vacuum wavenumbers and rotational assignments for the (1,0) and (2,0) bands of SH and SD are given in Tables I-IV. The relative accuracies for the measurements of the four bands are: SH (1,0), 0.1  $\text{cm}^{-1}$ ; SH (2,0), 0.3  $\text{cm}^{-1}$ ; SD (1,0), 0.05  $\text{cm}^{-1}$ ; SD (2,0), 0.2  $\text{cm}^{-1}$ .

TABLE II  
Vacuum wavenumbers and assignments of the rotational lines of the (2,0) SH band

$J$	$P_1$	$Q_1^0$ (and $Q_1^0 P_{21}$ )	$R_1$	$RQ_{21}$	$sR_{21}$	$P_2^0$ and $P_2^0 Q_{12}$	$Q_2^0$ (and $Q_2^0 R_{12}$ )
1.5	34,206.0	34,220.9	34,249.2*	34,249.2*	34,293.2	33,823.2*	33,852.7
2.5	175.2	203.9	246.9*	246.9*	304.5	803.5*	848.0
3.5	139.6	182.4	239.6*	239.6*	311.6	778.8	838.3
4.5	099.8	156.9	228.1*	228.1*			823.2*
5.5	056.3*	127.4	213.9	212.1			803.5*
6.5	008.7	094.1	194.1	192.2			
7.5	957.0	056.3*					
8.5		013.7					

\*Overlapped.

#### ANALYSIS

Rotational analyses of the SH and SD bands were carried out in the manner described previously (Ramsay 1952). The vibrational intervals  $\Delta G'_1$  and  $\Delta_2 G'_1$  were obtained from combination relations between corresponding lines of the (0,0) and (1,0) bands and of the (0,0) and (2,0) bands respectively. Thus it may readily be shown that

$$(2) \quad Q_1^{1,0}(J) - Q_1^{0,0}(J) = \Delta G'_1 - (B'_0 - B'_1)(J - \frac{1}{2})(J + \frac{1}{2}) + (D'_0 - D'_1)(J - \frac{1}{2})^2(J + \frac{1}{2})^2$$

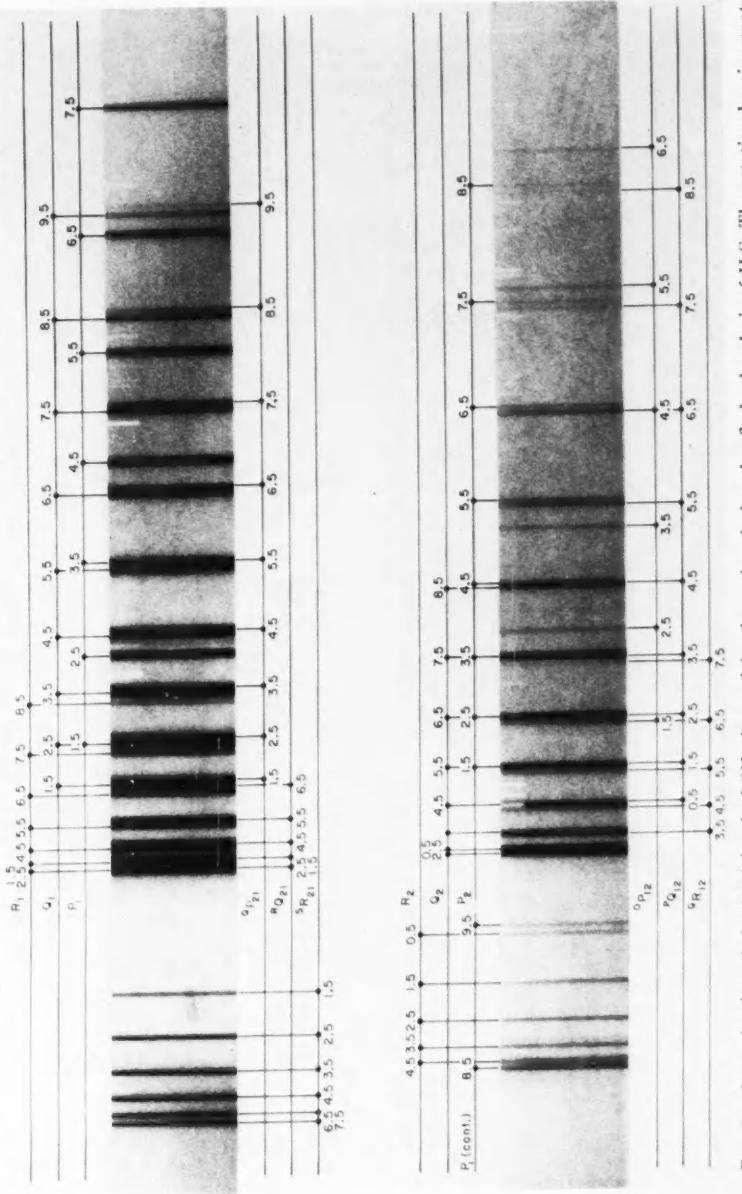


FIG. 1. Reproduction of the (1,0) band of SH observed in absorption during the flash photolysis of H<sub>2</sub>S. The rotational assignments are given. The R<sub>1</sub> head is at 3060.6 Å and the Q<sub>2</sub> head at 3098.3 Å.

PLATE II

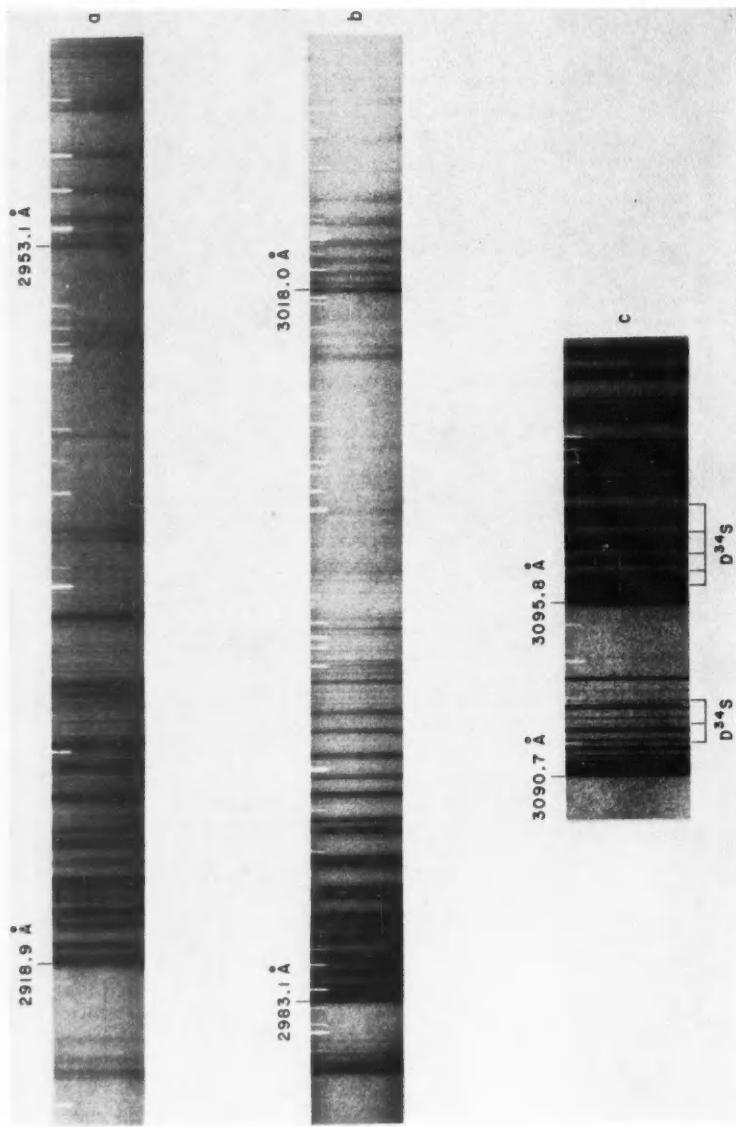


FIG. 2. Absorption bands of SH and SD observed during the flash photolysis of  $\text{H}_2\text{S}$  and  $\text{D}_2\text{S}$ ; (a) (2.0) band of  $\text{SH}$ , (b) (2.0) band of SD, (c) part of the (1.0) band of SD showing lines due to  $\text{D}^{34}\text{S}$  in natural abundance. The fine lines in (a) and (b) are due to absorption by  $\text{S}_2$  molecules.

TABLE III  
Vacuum wavenumbers and assignments of the (1,0) SD band

<i>J</i>	<i>P</i> <sub>1</sub>	<i>Q</i> <sub>1</sub>	<i>R</i> <sub>1</sub>	<i>Q</i> <sub>2</sub>	<i>R</i> <sub>2</sub>	<i>P</i> <sub>2</sub>	<i>Q</i> <sub>3</sub>	<i>R</i> <sub>3</sub>	<i>Q</i> <sub>4</sub>	<i>P</i> <sub>4</sub>	<i>Q</i> <sub>12</sub>
0.5	32,267.67†	32,275.79†	32,292.5†	32,275.79†	32,292.5†	32,292.5†	31,903.6†	31,920.0†	31,895.63*	31,903.6†	
2.5	251.68	268.28†	292.5†	267.97†	292.5†	325.25	880.51*	929.64	888.87*	905.4†	
3.5	234.42	259.19†	291.68*	258.62†	291.68*	332.46	870.43*	937.61	31,864.39	880.51*	905.4†
4.5	215.72	248.68†	289.84	248.00†	289.03	858.25	899.93	948.41	826.68	858.85*	
5.5	195.63	236.75	286.11	235.96	285.09	842.39	844.66	894.72	951.35*	805.19	845.52
6.5	174.27	223.39	280.91	222.34	279.69	845.15	829.40	887.82	952.22	782.10	830.48
7.5	151.10	208.56	274.19	207.33	272.84	846.16*	812.51	879.16	951.35*	757.34	888.87*
8.5	126.71	192.27	266.05	190.89	264.48	846.16*	793.94	868.79	949.11	813.72	880.51*
9.5	100.87	174.27	256.30	172.98	254.60	844.10	773.66	856.67	944.86	793.90	870.43*
10.5	073.57	155.18	245.00	153.53	243.18	840.55	751.62	842.82	938.81	775.16	858.85*
11.5	044.81	134.39	232.23	132.54	230.19	835.42	727.87	827.02	930.83	753.33	
12.5	014.57	112.06	217.80	110.05	215.72	702.70	809.63	809.63	729.74	704.51	
13.5	982.75	088.19	201.79	086.30	199.48	672.99	790.47				
14.5		062.73	185.18		181.66						
15.5		035.74									

\*Overlapped.  
†Barely overlapped.

‡Taken from earlier measurements.

TABLE IV  
Vacuum wavenumbers and assignments of the rotational lines of the (2,0) SD band

<i>J</i>	<i>P</i> <sub>1</sub>	<i>Q</i> <sub>1</sub> (and ${}^0 Q P_{21}$ )	<i>R</i> <sub>1</sub> and ${}^0 R_{21}$	<i>sR</i> <sub>21</sub>	<i>P</i> <sub>2</sub> and ${}^0 P_{12}$	<i>Q</i> <sub>2</sub> and ${}^0 Q_{12}$	<i>R</i> <sub>2</sub>
0.5						33,125.0*	33,140.3*
1.5	33,488.9*	33,497.3	33,512.1*	33,536.0	33,109.6	125.0*	148.5
2.5	472.9	488.9*	512.1*	543.3	100.3	125.0*	155.3
3.5	454.7	478.4	509.6	548.6	089.2	121.0	159.6
4.5	436.1*	466.3	505.1	552.2†	075.8	116.1	161.4*
5.5	413.4	452.6	498.4	552.2†	060.5	108.2	161.4*
6.5	389.7	436.1*	490.7	552.2†	042.5	098.6	
7.5	364.6	419.0	480.2	552.2†	022.7	086.0	
8.5	337.4	399.8		546.2		072.4	
9.5	308.4	378.3		540.2			
10.5	276.3	354.5					
11.5	245.4						
12.5	210.3						

\*Overlapped.

†Badly overlapped.

and that similar relations hold for the other 11 branches. By plotting the left-hand side of (2) against  $(J-\frac{1}{2})(J+\frac{1}{2})$  the points were found to lie on a definite curve. By extrapolating to  $(J-\frac{1}{2})(J+\frac{1}{2}) = 0$ , a value for  $\Delta G'_\frac{1}{2}$  was obtained. The quantity  $(\Delta G'_\frac{1}{2} - Q_1^{1,0}(J) + Q_1^{0,0}(J))/(J-\frac{1}{2})(J+\frac{1}{2})$  was then plotted against  $(J-\frac{1}{2})(J+\frac{1}{2})$  and a value for  $(B'_0 - B'_1)$  obtained from the intercept of the best straight line. The scatter of the points was such that it was not possible to determine an accurate value for the slope of the line and hence for  $(D'_0 - D'_1)$ . A value for  $\beta'$  ( $= D'_1 - D'_0$ ) was therefore calculated from the equation (Herzberg 1950, p. 108)

$$(3) \quad \beta' = D'_e \left( \frac{8\omega_e x_e}{\omega_e} - \frac{5\alpha'_e}{B'_e} - \frac{\alpha'_e \omega_e^2}{24B'_e^3} \right)$$

and was found to be consistent with the experimental data. As a final check on the constants  $\Delta G'_\frac{1}{2}$  and  $(B'_0 - B'_1)$ , the quantity  $Q_1^{1,0}(J) - Q_1^{0,0}(J) + \beta'(J-\frac{1}{2})^2(J+\frac{1}{2})^2$  was plotted against  $(J-\frac{1}{2})(J+\frac{1}{2})$ , together with similar relations for the other branches, and the best straight line was determined. The molecular constants obtained from the various bands are summarized as follows (in  $\text{cm}^{-1}$  units):

	SH	SD
$\Delta G'_\frac{1}{2}$	$1784.5 \pm 0.1$	$1319.30 \pm 0.05$
$B'_0 - B'_1$	$0.508 \pm 0.002$	$0.1828 \pm 0.0005$
$D'_1 - D'_0$ (calc.)	$0.6 \times 10^{-4}$	$0.1 \times 10^{-4}$
$\Delta_2 G'_1$	$3373.7 \pm 0.3$	$2540.9 \pm 0.2$
$B'_0 - B'_2$	$1.06 \pm 0.01$	$0.376 \pm 0.004$
$D'_2 - D'_0$ (calc.)	$1.2 \times 10^{-4}$	$0.2 \times 10^{-4}$

Values for the molecular constants  $B'_e$ ,  $\alpha'_e$ ,  $\gamma'_e$ ,  $D'_e$ ,  $\beta'_e$ ,  $r'_e$ ,  $\omega'_e$ , and  $\omega'_e x'_e$  were derived from the above data and the constants given earlier by Ramsay (1952); these are summarized in Table V. Attempts were made to obtain values for

TABLE V  
Summary of molecular constants for the  $A^2\Sigma^+$  state of  
SH and SD

	SH	SD	Units
$B_e$	8.521	4.392	$\text{cm}^{-1}$
$\alpha_e$	0.46 <sub>4</sub>	0.172	$\text{cm}^{-1}$
$\gamma_e$	-0.02 <sub>2</sub>	-0.005	$\text{cm}^{-1}$
$D_e$	$6.3_6 \times 10^{-4}$	$1.7_6 \times 10^{-4}$	$\text{cm}^{-1}$
$\beta_e$ (calc.)	$0.6_0 \times 10^{-4}$	$0.1_0 \times 10^{-4}$	$\text{cm}^{-1}$
$r_e$	1.423 <sub>0</sub>	1.423 <sub>5</sub>	$\text{\AA}$
$\omega_e$	1979.8	1417.0	$\text{cm}^{-1}$
$\omega_e x_e$	97.6 <sub>5</sub>	48.8 <sub>5</sub>	$\text{cm}^{-1}$

$\omega'_e y'_e$  by invoking isotope relations, but it was found that no significant values could be obtained.

#### DISSOCIATION ENERGY

An upper limit for the ground state dissociation energy of SH ( $D''_0 < 32,477 \text{ cm}^{-1}$ , 92.7 kcal/mole, 4.02 ev) has been established earlier (Ramsay 1952) from the predissociation observed in the (1,0) band. A lower limit may be established from the present work as follows. Since the  $v' = 2$  level of the  $^2\Sigma^+$  state lies below the dissociation limit H(<sup>2</sup>S) + S(<sup>1</sup>D), we have

$$(4) \quad \nu_0^{(2,0)} < D''_0 + \nu_8(^1D - ^3P)$$

i.e.  $D''_0 > 24,797 \text{ cm}^{-1}$  (70.9 kcal/mole, 3.07 ev).

From a linear extrapolation of the vibrational levels in the excited state, we obtain

$$(5) \quad D'_e = \omega_e'^2 / 4\omega'_e x'_e$$

Using the vibrational constants given in Table V, the values  $D'_e = 10,035 \text{ cm}^{-1}$  for SH and  $10,275 \text{ cm}^{-1}$  for SD are obtained. The two values are in satisfactory agreement and in subsequent calculations a mean value  $D'_e = 10,155 \text{ cm}^{-1}$  will be adopted. Now it is well known that for many molecules, though not for all, a linear extrapolation of the first few vibrational levels gives a value for the dissociation energy which is too high (Gaydon 1946). In particular for OH, for which the dissociation energy has been accurately determined by Barrow and Downie (1956), a linear extrapolation of the first few levels of the  $A^2\Sigma^+$  state gives a value for  $D'_e$  which is 21% too high. If we assume that a similar correction may be applied for SH, then  $D'_e = 8020 \text{ cm}^{-1}$ . From equation (1) we obtain  $D''_0 = 28,480 \text{ cm}^{-1}$  (81.4 kcal/mole, 3.53 ev) for SH and  $D''_0 = 28,850 \text{ cm}^{-1}$  (82.5 kcal/mole, 3.57<sub>5</sub> ev) for SD. Since the dissociation limit for the excited state lies only  $3700 \text{ cm}^{-1}$  above the highest observed vibrational level ( $v' = 2$  for SH) we consider that the values for the dissociation energies are probably accurate to  $\pm 1000 \text{ cm}^{-1}$  (2.9 kcal/mole, 0.12 ev).

## DISCUSSION

The intensity of absorption of the SH and SD bands in the present experiments is considerably greater than in the earlier work, the increase being due to the use of longer absorption paths and higher energies for the photolysis flash lamp. The intensity of absorption is now such that lines of the (1,0) band of the  $D^{34}S$  molecule in natural abundance (starting with  $D_2S$ ) are observed (see Fig. 2c). The isotope shifts between corresponding lines of  $D^{32}S$  and  $D^{34}S$  agree reasonably well with the value of  $1.7 \text{ cm}^{-1}$  calculated from the vibrational constants and the atomic masses. The corresponding lines of  $H^{34}S$  have not yet been detected presumably since the calculated shift is slightly smaller ( $1.1 \text{ cm}^{-1}$ ) and the lines of the (1,0) band are somewhat broader.

It is of interest to consider how many bands of the ( $v', 0$ ) progression may reasonably be expected to be observable in absorption. The Franck-Condon factors  $|\int \psi_{v'} \psi_{v''} dr|^2$  for SH and SD were evaluated by Drs. Fraser and Nicholls using the Fraser-Jarmain approximation method (1953); the results are summarized in Table VI. It is seen that the (0,0) bands are expected to be

TABLE VI  
Franck-Condon factors ( $|\int \psi_{v'} \psi_{v''} dr|^2$ )  
for SH and SD\*

$v'$	SH		SD
	$v'' = 0$	$v'' = 0$	$v'' = 0$
0	0.749		0.687
1	0.202		0.242
2	0.042		0.058
3	0.008		0.012
4	0.001		0.002

\*See also Nicholls, Fraser, Jarmain, and McEachran (1960).

from 2.8 to 3.7 times stronger than the (1,0) bands, which in turn should be approximately 4 to 5 times stronger than the (2,0) bands. These predictions are in qualitative agreement with observation. The (3,0) bands should be approximately 5 times weaker than the (2,0) bands and so far have not been detected.\* It may be noted that since lines of the (1,0) band of  $D^{34}S$  were observed weakly, and furthermore since the natural abundance of  $^{34}\text{S}$  is 4.2%, it should be possible to observe bands with Franck-Condon factors  $> 0.01$ . The (3,0) bands should therefore be on the limit of detection of the present apparatus. The failure to detect these bands\* may be due to the fact that the lines are probably quite broad due to predissociation. It would appear unlikely that any significant improvement in the determination of the dissociation energy of SH will be possible from further studies of the  $A \ ^2\Sigma^+ - X \ ^2\Pi$  system in absorption. It may be possible, however, to obtain a more precise value

\*Five broad weak lines of the (3,0) band of SD have now been observed in the region 2887-2890 Å.

spectroscopically from a study of other absorption systems at shorter wavelengths, or alternatively by studying suitable systems in emission as was carried out by Barrow and Downie for OH.

We wish to thank Drs. P. A. Fraser and R. W. Nicholls for carrying out the intensity calculations described above.

## REFERENCES

BARROW, R. F. and DOWNTON, A. R. 1956. Proc. Phys. Soc. (London), A, **69**, 178; see also  
BARROW, R. F. 1956. *Arkiv Fysik*, **11**, 281.  
CALLOMON, J. H. and RAMSAY, D. A. 1957. *Can. J. Phys.* **35**, 129.  
CLAESSON, S. and LINDQUIST, L. 1957. *Arkiv Kemi*, **11**, 535.  
EDLÉN, B. 1953. *J. Opt. Soc. Am.* **43**, 339.  
FRASER, P. A. and JARMAIN, W. R. 1953. Proc. Phys. Soc. (London), A, **66**, 1145, 1153.  
GAVDON, A. G. 1946. Proc. Phys. Soc. (London), A, **58**, 525.  
HARRISON, G. R. 1939. M.I.T. wavelength tables (John Wiley & Sons, Inc., New York).  
HERZBERG, G. 1950. Spectra of diatomic molecules (D. van Nostrand Co. Inc., Princeton,  
N.J.)  
NICHOLLS, R. W., FRASER, P. A., JARMAIN, W. R., and MCEACHRAN, R. P. 1960. *Astrophys.  
J.* **131**, 399.  
PORTER, G. 1950. *Discussions Faraday Soc.* **9**, 60.  
RAMSAY, D. A. 1952. *J. Chem. Phys.* **20**, 1920.  
WHITE, J. U. 1942. *J. Opt. Soc. Am.* **32**, 285.

---

## NOTES

---

### MACH'S PRINCIPLE AND SCALAR THEORIES OF GRAVITATION

PETER RASTALL

A number of theories of gravitation have recently been proposed in which the field is derived from a scalar potential (see references). In large part these theories can be written in generally covariant form, but they all contain some non-covariant element. There is always a preferred frame of reference in which the relation between the potential and the spacetime metric is particularly simple. It is the purpose of this note to point out that the lack of complete covariance is essential. It will be shown that any scalar theory of gravitation which is completely covariant must contradict Mach's principle.

Mach's principle is that the acceleration of a body in any frame of reference results entirely from its interactions with the universe (among which must be included its interactions with itself). The discovery of the correct interactions and the demonstration that the whole of a body's acceleration may be calculated from them constitute Mach's program. If for the purpose of the following argument we assume that Mach's principle is correct, then we can define an inertial frame of reference to be one in which the gravitational interactions of a small body do not cause it to accelerate. In other words, an inertial frame is one in which the gravitational field vanishes. It may be impossible to find any finite region throughout which the field vanishes exactly,\* but by going out into space, far from massive bodies, one would be able to find a region where it is negligible.

The discussion will be restricted to a class of theories in which the gravitational field is completely determined by a potential which is a scalar with respect to general transformations of co-ordinates. It will be assumed that a sufficient condition for the vanishing of the field in a spacetime region is that the potential should be constant in that region.

Consider a frame of reference in which the gravitational potential is constant in some spacetime region. In this region the field vanishes and, by Mach's principle, the frame is inertial. Take now a second frame of reference accelerating in any manner with respect to the first. Because the potential is an invariant with respect to general co-ordinate transformations, it is constant also in the corresponding region of the accelerated frame; the field vanishes there, and the accelerated frame is inertial. But it is known that frames accelerated with respect to a local inertial frame are not themselves inertial. We have proved a contradiction. Thus the class of scalar gravitational theories considered is incompatible with Mach's principle.

\*Even when the distribution of matter is perfectly uniform, many cosmological theories predict that freely falling bodies will be accelerated with respect to one another, the acceleration increasing with the separation of the bodies.

The possibility mentioned earlier, that the potential is not quite constant in the original frame of reference, does not alter the result of the argument. The deviations from constancy are exactly the same in the original and in the accelerated frame. One may easily arrange that the dimensions of the spacetime region considered be of the same order as measured in the two frames (it is only a question of keeping their relative speed much less than the speed of light). The field in the accelerated frame is then of the same order as in the other. If it is negligible in one it is negligible in both; and the accelerated frame is as close to being inertial as the original frame.

In conclusion it should be emphasized that, although a covariant scalar theory is impossible to reconcile with Mach's principle, yet a non-covariant one of the kind described in the references may be perfectly satisfactory. To prove its compatibility with Mach's principle it would be necessary to show that the preferred frames of reference are determined in some way by the matter distribution of the universe. This has not yet been done.

DICKE, R. H. 1957. *Rev. Modern Phys.* **29**, 363.  
PAPAPETROU, A. 1954. *Z. Physik*, **139**, 518.  
RASTALL, P. 1960. *Can. J. Phys.* **38**, 975.  
YILMAZ, H. 1958. *Phys. Rev.* **111**, 1417.

RECEIVED AUGUST 13, 1960.  
DEPARTMENT OF PHYSICS,  
UNIVERSITY OF BRITISH COLUMBIA,  
VANCOUVER 8, B.C.

#### ROCKET ELECTRON DENSITY MEASUREMENTS AT FORT CHURCHILL, CANADA

A. W. ADEY AND W. J. HEIKKILA

The Defence Research Telecommunications Establishment has instrumented an Aerobee-150 rocket (DRTE 01) for electron density measurements using the two-frequency, phase-comparison method of Seddon (1953) and Chapman *et al.* (1960). The rocket was fired during quiet ionospheric conditions at 1237 C.S.T. on 17 September, 1959, at Fort Churchill, Canada.

In the experiment the phases of two harmonically related signals from the rocket during flight are compared on the ground, and produce beat notes which are recorded. A different beat note is produced by each of the ordinary and extraordinary waves.

The two transmissions from the rocket were made at frequencies of 7.775 and 46.65 Mc/s respectively. At a second ground station the same antennas were arranged to receive linearly polarized signals so that the apparent spin rate of the rocket could be measured. The difference between this apparent spin rate and the true spin rate as determined by means of a solar aspect cell provided a measure of the Faraday rotation effect on the transmissions, and from this a further calculation of electron density was made.

The results obtained are presented in Fig. 1. The altitude scale may be in error by 1 or 2%, because of inaccurate tracking data. The Faraday rotation data are shown only for the lowest portion of the ascending part of the flight, because at higher altitudes and more oblique propagation angles multipath

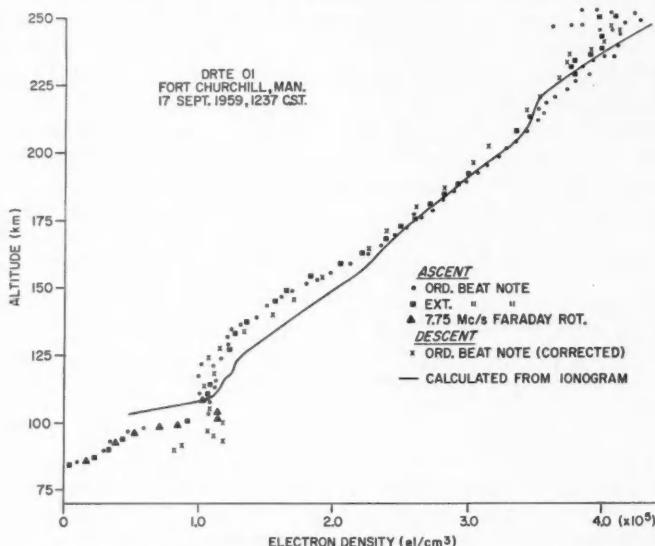


FIG. 1. The electron density profile obtained from the rocket measurements.

interference made this record unreliable. The two sets of phase-comparison data for the ascent leg differ slightly due to the neglect, in the calculation, of the variation with altitude of the magnetic field strength and the angle between the magnetic field lines and the direction of propagation.

For the descent part of the flight only the ordinary beat note record was usable. The data based on this record were corrected for obliquity effects by use of the method of Jackson (1957). This correction for the ascent data was found to be negligible.

The agreement between the ascent and descent data is close, except for the extremes of the altitude range considered. Application of a correction for ray bending (Jackson 1957) brought the two sets of data into closer agreement at the higher altitudes but did not affect the lower-altitude data. However, a sporadic-*E* condition existed during the flight, and this may have accounted for the lower-altitude discrepancy. In addition, because of the rocket's horizontal component of velocity, approximately 100 km separated the regions of the *E*-layer traversed by the rocket on the upward and downward parts of the flight.

During the flight the Canadian Department of Transport operated an ionospheric sounder continuously at a station about 12 km from the launcher.

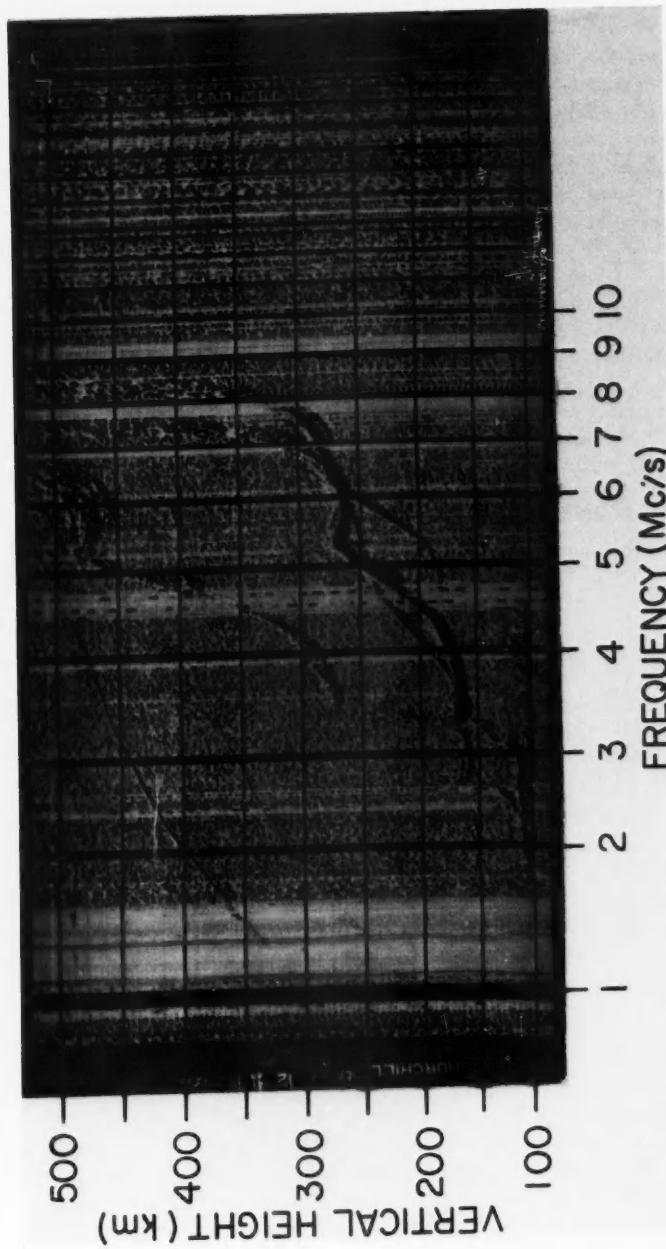
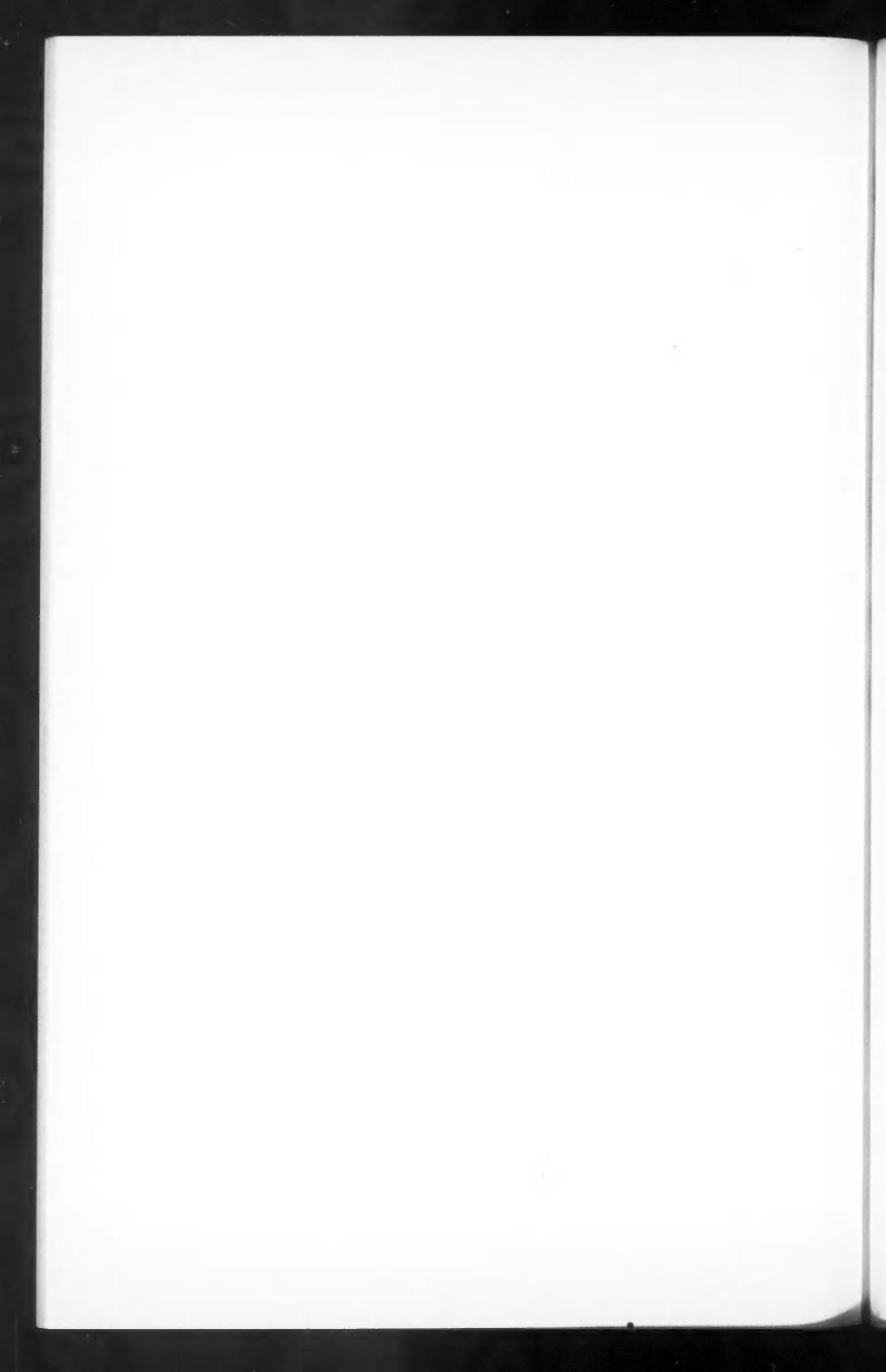


FIG. 2. Ionogram obtained at the Department of Transport sounding station at Fort Churchill during the flight of DRTE rocket #01.



Slight changes were noticed during the flight, particularly in *F*-region structure. One of the ionograms is reproduced in Fig. 2. A true height analysis (Wright and Norton 1959) of this ionogram has yielded a profile in good agreement with the rocket results, as shown in Fig. 1, but only after very careful scaling of the *E*-region traces. In particular, the subsidiary stratification shown near 130 km by the rocket results produces an apparent critical frequency of 3.3 Mc/s, while the *E*-region critical frequency is just over 3.0 Mc/s. The weak sporadic-*E* trace that extends up to 4.5 Mc/s may be related to the small cusp in the rocket profile at 100 km, or to the steep gradient just below 100 km.

We wish to thank Mr. S. R. Penstone for the electrical design of the nose cone, and Mr. D. Burke for the mechanical design and trajectory calculations. We also wish to thank Mr. E. Moskal for his calculations of the results for the descent portion of the flight, and Mr. R. Maliphant for the reduction of the ionogram. We acknowledge the assistance of J. E. Jackson (formerly of N.R.L. and now of N.A.S.A.) and his staff in the design of the transmitter. They also provided, and helped to operate, the ground station.

CHAPMAN, J. H., HEIKKILA, W. J., and BROWN, R. K. 1960. Two High Altitude Rocket Experiments, Proc. Cospar Space Symposium, Jan. 1960.  
 JACKSON, J. E. 1957. Effect of Oblique Propagation Paths Upon the NRL Rocket Studies of the Ionosphere, NRL Report 4960.  
 SEDDON, J. C. 1953. J. Geophys. Research, **58**, 323.  
 WRIGHT, J. W. and NORTON, R. B. 1959. N.B.S. Report 6031.

RECEIVED OCTOBER 24, 1960.

DEFENCE RESEARCH TELECOMMUNICATIONS ESTABLISHMENT,  
 DEPARTMENT OF NATIONAL DEFENCE,  
 SHIRLEY BAY, ONT.

#### THE SPECIFIC FLUORESCENCE OF PLASTIC SCINTILLATOR NE102

J. R. PRESCOTT AND A. S. RUPAAL

##### 1. INTRODUCTION

Studies of crystalline organic materials (Birks 1951, 1952, 1953; Birks and Brooks 1956; Fowler and Roos 1955; Swartz *et al.* 1957; Taylor *et al.* 1951) have shown that the specific fluorescence  $dS/dx$  in response to the passage of ionizing particles can be well represented by the formula of Birks (1951):

$$(1) \quad \frac{dS}{dx} = \frac{A(dE/dx)}{1+kB(dE/dx)}$$

where  $dE/dx$  is the specific energy loss and  $A$  and  $kB$  are constants for the material in question. The numerator represents the production of excited molecules and the second term in the denominator, the quenching of the light output by damaged molecules along the track of the particle,  $B(dE/dx)$  being the number of damaged molecules per undamaged molecule and  $k$  the relative probability that quenching rather than emission will occur.

Birks' formula has also been used to describe the response of plastic scintillators (Boreli and Grimeland 1955; Daehnick and Fowler 1958; Evans and Bellamy 1959). A value for  $kB$  of  $0.010 \pm 0.001 \text{ g cm}^{-2} \text{ Mev}^{-1}$  in NE102 plastic scintillator which has a polyvinyl-toluene base, has been obtained by Evans and Bellamy (1959), using protons of energy 1–14 Mev. This value is consistent with the work of Gettner and Selope (1960), who used recoil protons from neutrons in the energy range 0.14 to 2.3 Mev. Daehnick and Fowler (1958), on the other hand, report values of 0.002 and  $0.003 \pm 0.001$  for  $kB$ . Boreli and Grimeland (1955) found a value of  $0.011 \text{ g cm}^{-2} \text{ Mev}^{-1}$  for a polystyrene base plastic.

In the present work, values for  $kB$  have been obtained for a number of different samples of plastic scintillator NE102\* by comparing the response of the samples to gamma rays and alpha particles, and for one sample the result has been checked using recoil protons.

Since gamma-ray detection in a plastic scintillator takes place by the production of Compton electrons,  $dE/dx$  is small for sufficiently large  $E$  and under these conditions Birks' formula (1) becomes:

$$(2) \quad dS_e/dE = A$$

where  $A$  can be expressed in any convenient units.

For an alpha particle,  $dE/dx$  is sufficiently large that saturation occurs along the track and (1) becomes:

$$(3) \quad dS_a/dx = A/kB.$$

Since the alpha particles are incident on the crystal from outside, expression (3) must be modified to account for surface effects which are important when the residual range of the particle is less than about 1 cm air equivalent. This effect has been discussed by Birks and Brooks (1956) and interpreted in terms of escape and/or quenching of fluorescence excitation energy at the surface. For residual ranges less than about 1 cm air equivalent, the right-hand side of equation (3) is multiplied by a function that varies from 0.5 to 1; it remains valid, however, for residual ranges greater than about 1 cm air equivalent. The effect is not important for the gamma rays since the electrons are produced within the body of the scintillator itself. However, the presence of the  $kB$  term in (1) introduces a small non linearity for electron energies less than about 100 kev.

The value of  $kB$  may be determined by measurements with gamma rays and alpha particles in the same sample under similar conditions. Combining (2) and (3):

$$(4) \quad kB = \frac{dS_e}{dE} / \frac{dS_a}{dx}.$$

## 2. EXPERIMENTAL

Polished cylindrical samples of NE102, 0.7 cm thick, 4 cm in diameter, were mounted on an RCA 6342 photomultiplier and the pulse height spectra dis-

\*Nuclear Enterprises, 1750 Pembina Highway, Winnipeg, Canada. Dated samples from different batches of scintillator were kindly supplied by Nuclear Enterprises.

played on a CDC 100-channel kicksorter. The samples were surrounded by a polished aluminum reflector. A hole in the reflector, 0.5 cm in diameter, admitted alpha particles to the crystal surface. Pulse height spectra were taken as a function of distance from the sample surface of a thin source of 5.3-Mev alpha particles. A typical curve of pulse height vs. residual range in air (reduced to 760 mm and 15°C, Jesse and Sadauskis 1950) is shown in Fig. 1. It is seen

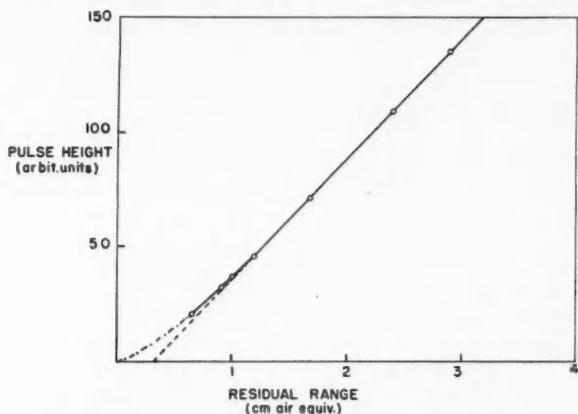


FIG. 1. Pulse height as a function of residual range for alpha particles incident on NE102 plastic scintillator.

that the pulse height is a linear function of residual range for ranges greater than about 1 cm air equivalent. If the non linearity for small ranges is interpreted as due to escape of excitation energy from the crystal, it implies an absorption length for this energy of about 5 mm air equivalent which may be compared with ranges of 3-4 mm air equivalent for anthracene, stilbene, and terphenyl found by Birks (1952). The measurements are not accurate enough to establish whether the figure for NE102 is significantly greater than that for the crystalline compounds.

The gamma-ray response was determined with gamma rays of energy 279, 511, and 662 kev. The Compton edge was taken at  $\frac{2}{3}$  of the maximum height of the Compton peak. It appears to be common practice to take the Compton edge at half height. However, a family of curves for which various Gaussian resolution functions were folded into theoretical Compton spectra showed that  $\frac{2}{3}$  height coincides much better with the Compton edge (within about 1%) for a very wide range of gamma-ray energies and resolutions. Graphs of pulse height vs. gamma-ray energy were linear (as expected) but had an intercept on the energy axis of about 10 kev.

Using relation (4) the values of  $kB$  shown in Table I were obtained. The table also shows relative values of  $A$ . With one exception, sample 159, the  $A$  values do not differ significantly. The first five entries are identified by manufacturer's batch number and date of manufacture. The last three were samples cut from the same batch but having different histories. Sample UBC 1

TABLE I

*kB* values and relative *A* values in the formula

$$\frac{dS}{dx} = \frac{A(dE/dx)}{1+kB(dE/dx)} \text{ for plastic NE102}$$

Batch No.	Date of manufacture	<i>kB</i> cm air equiv. Mev <sup>-1</sup> ± 4%	Relative <i>A</i> ± 3%
159	March 1, 1958	9.6	1.0
176	July 7, 1958	8.9	0.86
257	Feb. 12, 1959	9.8	0.82
269	June 15, 1959	8.6	0.89
294	April 4, 1960	8.8	0.87
UBC 1	Early 1958	9.0	0.88
UBC 2	Early 1958	9.8	0.82
UBC 3	Early 1958	12.9	0.85

had been maintained almost continuously in the dark since manufacture, sample UBC 2 had been exposed intermittently to light over a period of  $2\frac{1}{2}$  years; sample UBC 3 had been exposed almost continuously to light and at times to strong sunlight for an undetermined period. It is clear that, although its relative response appears unchanged, sample UBC 3 has a significantly larger value of *kB*. Since *kB* is a measure of damaged molecules, it is to be supposed that exposure to strong light has resulted in significant damage to the scintillator. Excluding the data for this sample, the average value of *kB* for all samples is  $9.21 \pm 0.17$  cm air equivalent Mev<sup>-1</sup>.

This figure may be converted to cm plastic Mev<sup>-1</sup> by calculating the range of alpha particles in the plastic. This was done by using the method of Livesey (1956). In Livesey's method, the specific energy loss as a function of velocity is represented by a power law for velocities much greater than the orbital velocity of the first electron in the *K* shell of the ion, departure from the power law at low ion velocities being attributed to a decreasing effective charge on the ion. The specific energy loss of protons in NE102 was computed as a function of velocity by combining the data for carbon and hydrogen given by Whaling (1958). The effective charge of a proton at low velocities was deduced from this curve using Livesey's approximation and found to be in close agreement with the effective charge for protons in air given by Stier *et al.* (1954) and Allison and Warshaw (1953). Using effective charges for alpha particles obtained by averaging the data of Nikolaev *et al.* (1957), Whaling (1958, Table 2d), Stier *et al.* (1954), and Snitzer (1953), the specific energy loss of alpha particles was deduced and the range as a function of energy found by numerical integration in the usual way. Finally, comparison with the range-energy data for alpha particles in air (Jesse and Sadauskis 1950) showed that 1 cm air is equivalent to  $1.03 \times 10^{-3}$  ( $\pm 7\%$ ) cm NE102. The estimated 7% error includes the uncertainties in the stopping power, effective charge, and the additive integration error due to the uncertainty of the behavior of the ion at low velocities.

The value of  $kB$  may therefore be quoted as

$$\begin{aligned} & 9.21 \pm 0.17 \text{ cm air equivalent Mev}^{-1} \\ & = 0.0095 \pm 0.0008 \text{ cm plastic Mev}^{-1} \\ & = 0.0098 \pm 0.0008 \text{ g cm}^{-2} \text{ Mev}^{-1}. \end{aligned}$$

A check on this figure was provided by a somewhat different method. Sample UBC 2 was exposed to collimated fluxes of monochromatic neutrons from the D-D reaction at energies from 2 to 4 Mev. In a scintillator sample of this size and at these energies, the probability of multiple neutron collisions is negligible and the differential energy spectrum of recoil protons from neutron collisions is rectangular from zero energy to the full neutron energy. The integral spectrum of such recoils should then be a straight line that (apart from the effect of finite photomultiplier resolution) extrapolates to cut the energy axis at the full neutron energy. In this proton energy range, however,  $dE/dx$  in relation (1) can neither be regarded as small nor large and the pulse height is a markedly non-linear function of proton energy. Because of this non-linear luminescence response, the observed pulse height spectrum is distorted and the integral curve is concave upwards, an excess of small pulses and a defect of large pulses being observed. The curved integral plot can be converted into a straight line by distorting the scale on the energy axis appropriately, using equation (1) (Swartz *et al.* 1957). Values of  $dE/dx$  for protons in NE102 were obtained as described earlier in this paper and equation (1) integrated numerically to give  $S$  as a function of  $E$  with  $kB$  as a parameter. The constant  $A$  was found from a gamma-ray calibration. Differential neutron spectra were obtained at seven neutron energies in the region 2 to 4 Mev. From these, integral spectra were plotted. A value of  $kB$  was then sought by trial and error that: (a) converted the integral plot into a straight line (b) yielded the correct primary neutron energy on extrapolation. A typical plot before and after transformation is shown in Fig. 2. The conversion to a straight line is seen to be good except at low energies where the background correction becomes appreciable and at high energies where photomultiplier statistics produce a tail. The average value of  $kB$  for the seven neutron energies was  $0.0088 \pm 0.0002$  cm plastic Mev $^{-1}$  (sample standard error). Allowing 5% uncertainty for the specific energy loss figures and the integration,

$$\begin{aligned} kB & = 0.0088 \pm 0.0006 \text{ cm plastic Mev}^{-1} \\ & = 0.0091 \pm 0.0006 \text{ g cm}^{-2} \text{ Mev}^{-1}. \end{aligned}$$

Although, statistically speaking, this figure does not differ significantly from the value found using alpha particles, the two values may, in fact, be expected to differ. In the recoil proton method of measuring  $kB$ , the light is produced by protons and gamma rays within the body of the crystal which is completely surrounded by reflector. The optical geometry is thus substantially the same for both. In the alpha-particle method, however, the light from the alpha particles is produced very close to the surface of the crystal in the region of

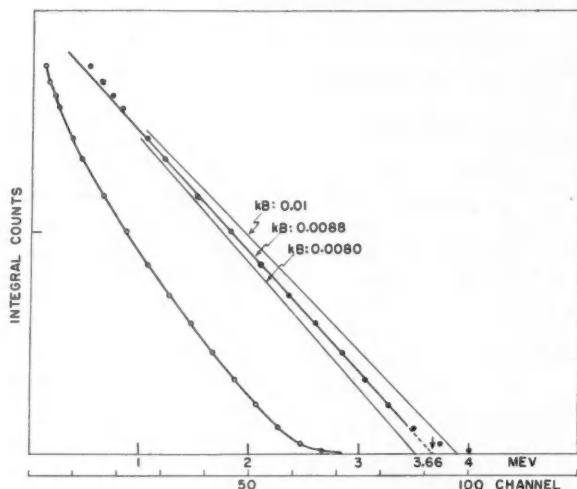


FIG. 2. Integral pulse height spectra of recoil proton pulses from 3.66-Mev neutrons in NE102 plastic scintillator. Open circles: observed spectrum. Solid circles: spectrum corrected for non-linear luminescence response using  $kB = 0.0088 \text{ cm Mev}^{-1}$ . Lines for  $kB = 0.010$  and  $0.0080$  are also shown for comparison.

the window in the reflector. Some light from the alpha-particle pulse may therefore escape through the surface of the crystal without being returned by the reflector. The gamma-ray calibration under the same conditions produces light throughout the thickness of the crystal, and a relatively smaller fraction can escape through the window in the reflector. The effect of this is to depress the height of alpha-particle pulses with respect to gamma-ray pulses, leading to a high value of  $kB$ . In the present experiments the gamma rays were collimated into the crystal at the same point as the alpha particles to try to compensate for this effect but a rough calculation suggests that the  $kB$  values for the alpha-particle method might be as much as 10% too high, which is of about the magnitude of the difference observed. Support for this view was provided by an experiment in which the pulse heights for alpha-particle and gamma-ray pulses were measured with and without reflector. The reduction in pulse height for the gamma rays on removal of the reflector was 7% larger than that for the alpha particles and, although not sufficient to account quantitatively for the difference between the two  $kB$  values, is in the right direction.

For this reason we believe that the value of  $kB$  for NE102 is closer to the value  $0.0091 \pm 0.0006 \text{ g cm}^{-2} \text{ Mev}^{-1}$  obtained using recoil protons than to the figure  $0.0098 \pm 0.0008 \text{ g cm}^{-2} \text{ Mev}^{-1}$  obtained using alpha particles. Either of these figures is in agreement with the determination of Evans and Bellamy (1959), but the low values obtained by Daehnick and Fowler (1958) remain unexplained. Possible variability from batch to batch of scintillator would appear to be ruled out by the data displayed in Table I.

ALLISON, S. K. and WARSHAW, S. D. 1953. *Rev. Modern Phys.* **25**, 779.  
BIRKS, J. B. 1951. *Proc. Phys. Soc. (London)*, A, **64**, 874.  
— 1952. *Phys. Rev.* **86**, 569.  
— 1953. Scintillation counters (Pergamon Press, London), p. 93.  
BIRKS, J. B. and BROOKS, F. D. 1956. *Proc. Phys. Soc. (London)*, B, **69**, 721.  
BORELLI, F. and GRIMELAND, B. 1955. *Nuovo Cimento*, **2**, 336.  
DAEHNICK, W. W. and FOWLER, J. M. 1958. *Phys. Rev.* **111**, 1309.  
EVANS, H. C. and BELLAMY, E. H. 1959. *Proc. Phys. Soc. (London)*, **74**, 483.  
FOWLER, J. M. and ROOS, C. E. 1955. *Phys. Rev.* **98**, 996.  
GETTNER, M. and SELOVE, W. 1960. *Rev. Sci. Instr.* **31**, 450.  
JESSE, W. P. and SADAUSKIS, J. 1950. *Phys. Rev.* **78**, 1.  
LIVESEY, D. L. 1956. *Can. J. Phys.* **34**, 203.  
NIKOLAEV, V. C., DMITRIEV, I. C., FATEEVA, L. N., and TEPLOVA, Y. A. 1957. *J. Exptl. Theoret. Phys. U.S.S.R.* **33**, 1325. Translation *J.E.T.P.* **6**, 1019, 1958.  
SNITZER, E. 1953. *Phys. Rev.* **89**, 1237.  
STIER, P. M., BARNETT, C. F., and EVANS, G. E. 1954. *Phys. Rev.* **96**, 973.  
SWARTZ, C. D., OWEN, G. E., and AMES, O. 1957. "The Stilbene Crystal as a Neutron Spectrometer", NYO-2053.  
TAYLOR, C. J., JENTSCH, W. K., REMLEY, M. E., EBY, F. S., and KRUGER, P. G. 1951. *Phys. Rev.* **84**, 1034.  
WHALING, W. 1958. *Handbuch der Physik* Vol. 34 (Springer-Verlag, Berlin), p. 193.

RECEIVED SEPTEMBER 29, 1960.  
PHYSICS DEPARTMENT,  
UNIVERSITY OF BRITISH COLUMBIA,  
VANCOUVER, B.C.

## LETTERS TO THE EDITOR

Under this heading brief reports of important discoveries in physics may be published. These reports should not exceed 600 words and, for any issue, should be submitted not later than six weeks previous to the first day of the month of issue. No proof will be sent to the authors.

### Frequency Measurement of Standard Frequency Transmissions<sup>1,2</sup>

Measurements are made at Ottawa, Canada, using N.R.C. caesium-beam frequency resonator as reference standard (with an assumed frequency of 9 192 631 770 c.p.s.). Frequency deviations from nominal are quoted in parts per  $10^{10}$ . A negative sign indicates that the frequency is below nominal.

Date, October 1960	GBR, 16 kc/s			
	MSF, 60 kc/s	7-hour average*	24-hour average	WWVB, 60 kc/s
1	-160	-156	-157	N.M.
2	N.M.	-159	-156	N.M.
3	-167	-161	-156	-145
4	N.M.	-159	-155	-146
5	N.M.	N.M.	N.M.	N.M.
6	N.M.	N.M.	N.M.	N.M.
7	N.M.	N.M.	N.M.	N.M.
8	N.M.	N.M.	N.M.	N.M.
9	N.M.	N.M.	N.M.	N.M.
10	N.M.	N.M.	N.M.	N.M.
11	N.M.	-160	-154	-142
12	-160	-162	-163	-145
13	-149	-162	-156	-145
14	-152	-160	-157	-146
15	-161	-157	-158	N.M.
16	-158	-158	-157	N.M.
17	-157	-155	-154	-146
18	-161	-165	-157	-148
19	-156	-158	-154	-145
20	-155	-158	-157	-145
21	-166	-161	-158	-145
22	-153	-160	-157	N.M.
23	-153	-164	-160	N.M.
24	-155	-166	-159	-147
25	-171	-162	-160	-146
26	-166	-166	-160	-148
27	N.M.	-167	-161	-149
28	-162	N.M.	-161	-148
29	N.M.	N.M.	N.M.	N.M.
30	N.M.	N.M.	N.M.	N.M.
31	-167	N.M.	N.M.	-150
Midmonthly mean	-159	-161	-158	-146
Midmonthly mean of WWV	-144			

NOTE: N.M. no measurement.

\*Time of observations: 00.00 to 05.30 and 22.30 to 24.00 U.T.

RECEIVED NOVEMBER 17, 1960.  
DIVISION OF APPLIED PHYSICS,  
NATIONAL RESEARCH COUNCIL,  
OTTAWA, CANADA.

S. N. KALRA

<sup>1</sup>Issued as N.R.C. No. 6084.

<sup>2</sup>Cf. Kalra, S. N. 1959. Can. J. Phys. 37, 1328.

### The Sudden Increase in the Photon Component of Cosmic Radiation on May 4, 1960

Results obtained at four locations in Canada with both neutron and meson monitors during the solar flare of May 4, 1960, have recently been published (Rose 1960). The purpose of this letter is to report the observations made at a fifth location (Winnipeg, Man., geographic latitude  $49.9^{\circ}$  N., longitude  $97.2^{\circ}$  W., and altitude 236 meters) with a cosmic-ray monitor designed primarily to study the photon component of the cosmic radiation.

The monitor used in these experiments is a large crystal total absorption spectrometer which has previously been described (Standil and Loveridge 1959). Essentially it is a large NaI(Tl) crystal (over-all dimensions  $\sim 9$  in.) which is completely surrounded by an efficient scintillating plastic anticoincidence shield. Events in the NaI crystal unaccompanied by events in the shield are thus initiated by neutral radiation, primarily the photon component.

We have been simultaneously recording the integral counting rates from our NaI crystal with and without the anticoincidence shield at two integral discriminator settings, namely 12 and 35 Mev. For each of these energy levels we therefore obtain the counting rates for the detected neutral radiation and (by subtraction of the two recorded rates) for the events due to charged and neutral but not completely absorbed particles. Calling these respectively the P and C events, the former are primarily due to the photon component and the latter primarily due to  $\mu$ -mesons. Our method of recording involves four scalers which are photographed every 15 minutes. The time resolution is therefore quite poor but this represents a compromise between the desirability of both reasonable counting statistics and time resolution. Our observed results are given in Fig. 1. The counting rate is plotted as a percentage of the mean

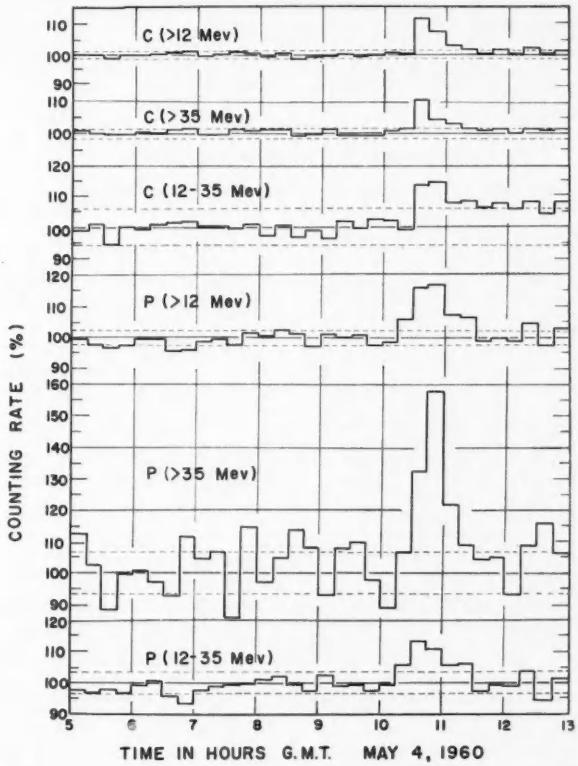


FIG. 1. Counting rates (uncorrected for pressure changes) as a percentage of the average rates from 0500 to 1300 hours May 4, 1960. The curves labelled C refer essentially to the  $\mu$ -meson component while the curves labelled P are due primarily to the photon component. The energies mentioned are the energies lost in the NaI detector by the particles concerned. Only in the case of the P curves are these the total energies of the particles (i.e. photons) concerned.

counting rate over the period from 0500 to 1300 hours and in each case the horizontal dashed lines represent the limits of the probable counting error. As the pressure remained fairly constant over this period no corrections have been made for pressure variations. The following preliminary comments may be made about these results:

(1) The time at which the increases occur,  $10:45 \pm 15$  minutes, agrees with the times of increase noted by Rose (1960).

(2) The magnitude of the increase for the C component at Winnipeg (essentially  $\mu$ -mesons), 10–15%, is in general agreement with the reported meson telescope data for Churchill, Sulphur Mountain, and Ottawa.

(3) The magnitude of the increases for the P component (essentially  $\gamma$ -rays) is much smaller than the increases reported at the above three locations for the neutron component, the latter ranging from about 300 to 430%. Moreover with the energy discrimination available in this experiment it is interesting to note that the increase for those events above 35 Mev (~50–60%) is considerably greater than the increase for those events in the lower energy range from 12–35 Mev (~10–15%). The meaning of this may become clear with the collection of further data.

This work was supported by the National Research Council of Canada and the Research Corporation.

ROSE, D. C. 1960. Can. J. Phys. **38**, 1224.  
STANDIL, S. and LOVERIDGE, W. D. 1959. Rev. Sci. Instr. **30**, 931.

RECEIVED NOVEMBER 18, 1960.  
DEPARTMENT OF PHYSICS,  
UNIVERSITY OF MANITOBA,  
WINNIPEG, MAN.

S. STANDIL  
R. P. BUKATA  
F. K. CHIN

### The Nuclear Spin of Indium-115<sup>1</sup>

The nuclear spin of the 4.5-hour metastable state of In<sup>115</sup> has been measured using the "flop-in" method of atomic beam magnetic resonance (Zacharias 1942). Details of the construction of the atomic beam apparatus will be given at a later date.

The indium isotope was produced in the McMaster reactor by neutron irradiation of cadmium foil. The cross section for thermal neutron capture in Cd<sup>114</sup> (natural abundance ~29%) is approximately 1 barn. The resulting Cd<sup>115</sup> decays with a 54-hour half-life into the desired metastable state of indium. Irradiation of 2-gram samples in a flux of  $10^{13}$  neutrons/cm<sup>2</sup> sec for 8 hours a day for a week resulted in about 100 millicuries of activity.

Chemical separation of the indium was carried out 15 hours after the end of irradiation. Fifteen milligrams of indium carrier were added to the concentrated HCl-HNO<sub>3</sub> mixture in which the cadmium was dissolved. In an excess of ammonia, cadmium forms a soluble complex but the indium precipitates as hydroxide. After precipitation, the In(OH)<sub>3</sub> was separated and dissolved in weak hydrochloric acid. Metallic indium was easily recovered from the acid solution by reduction with powdered magnesium. It was subsequently packed in a carbon crucible which fitted inside the tantalum oven of the atomic beam apparatus. A satisfactory beam of indium atoms, lasting 3 or 4 hours, was obtained at an oven temperature of 1100°C.

The radioactive indium beam was detected by deposition on collector buttons placed at the exit slit of the magnet system. In order to improve the collection efficiency, the buttons were plated with copper immediately prior to use. After exposure to the beam, the buttons were removed from the vacuum and examined for radioactivity using  $2\pi$ -geometry scintillation spectrometers. The windows of the single-channel pulse height analyzers were set to accept pulses corresponding to the photopeak of 335-kev gamma rays. Ninety-five per cent of In<sup>115m</sup> de-excites with the emission of such radiation.

The magnetic field (C field) in which the r-f. transitions were effected was calibrated using the resonance in a caesium beam. This was obtained from an alternate oven at lower temperature and detected using a surface ionization detector which could be placed at the exit slit instead of a button. This detector could also be used to monitor the over-all indium beam but, of course, it does not differentiate between radioactive and non-radioactive atoms.

Spin searches, in which series of buttons were exposed at selected radio-frequencies, were conducted at several low values of C field. At each magnetic field a single low-frequency resonance,  $\Delta F = 0$ ,  $\Delta m_F = \pm 1$ , was found corresponding to  $J = \frac{1}{2}$ ,  $I = \frac{1}{2}$ . The result of one of these runs is shown in Fig. 1. No serious attempt was made to observe resonances in the low-lying  $^2P_{3/2}$  atomic state of indium. The activity on the  $I = \frac{1}{2}$  buttons was followed for 12 hours and found to decay with a half-life of  $4.4 \pm 0.4$  hours. This fact, together with con-

<sup>1</sup>This work is based on the Ph.D. thesis of one of the authors (H.J.K.).

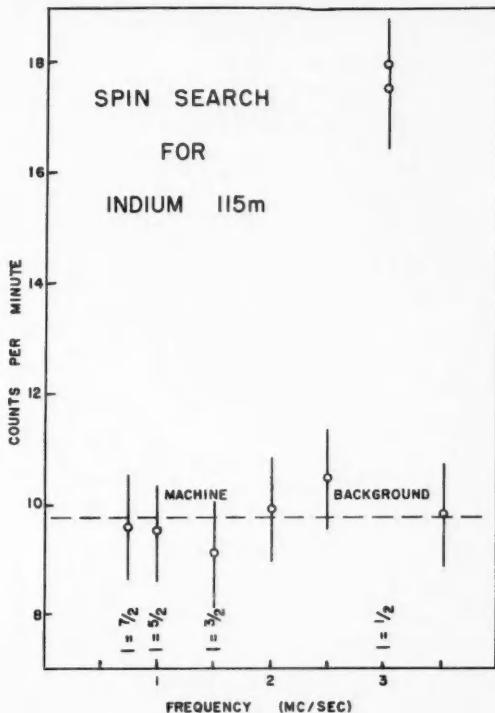


FIG. 1. Results of a spin search at a  $C$  field of 6.4 gauss. The experimental counting rates have been corrected for decay and for variations in beam intensity. Counter background has been subtracted but "machine background" due to gas scattering, etc. remains.

sideration of the mode of production of the radioactivity and of the method of radioactive detection, definitely establishes that  $I = \frac{1}{2}$  for  $\text{In}^{115m}$ .

This conclusion is in agreement with the assignment  $M4$  (Varma and Mandeville 1955) for the gamma-ray transition between the metastable state and the  $I = 9/2$  ground state (Millman *et al.* 1938). According to the shell model the metastable state arises when the hole in the  $g_{9/2}$  proton level is filled leaving a hole in the  $p_{1/2}$  level instead.

Work is being continued to measure the spins of the 1.1-hour ground state and the 1.9-hour metastable state of  $\text{In}^{117}$ . A determination of the magnetic moments of these isotopes and of  $\text{In}^{115m}$  is planned.

R. H. Tomlinson and K. Fritze of the Department of Chemistry have given valuable help in developing the chemical procedure.

The authors wish to acknowledge the generous financial support of this project by the National Research Council. H. J. King is grateful for the award of N.R.C. studentships in the period 1957-60. The Ontario Research Foundation has given substantial support in the form of scholarships (1958-60) to J. A. Cameron.

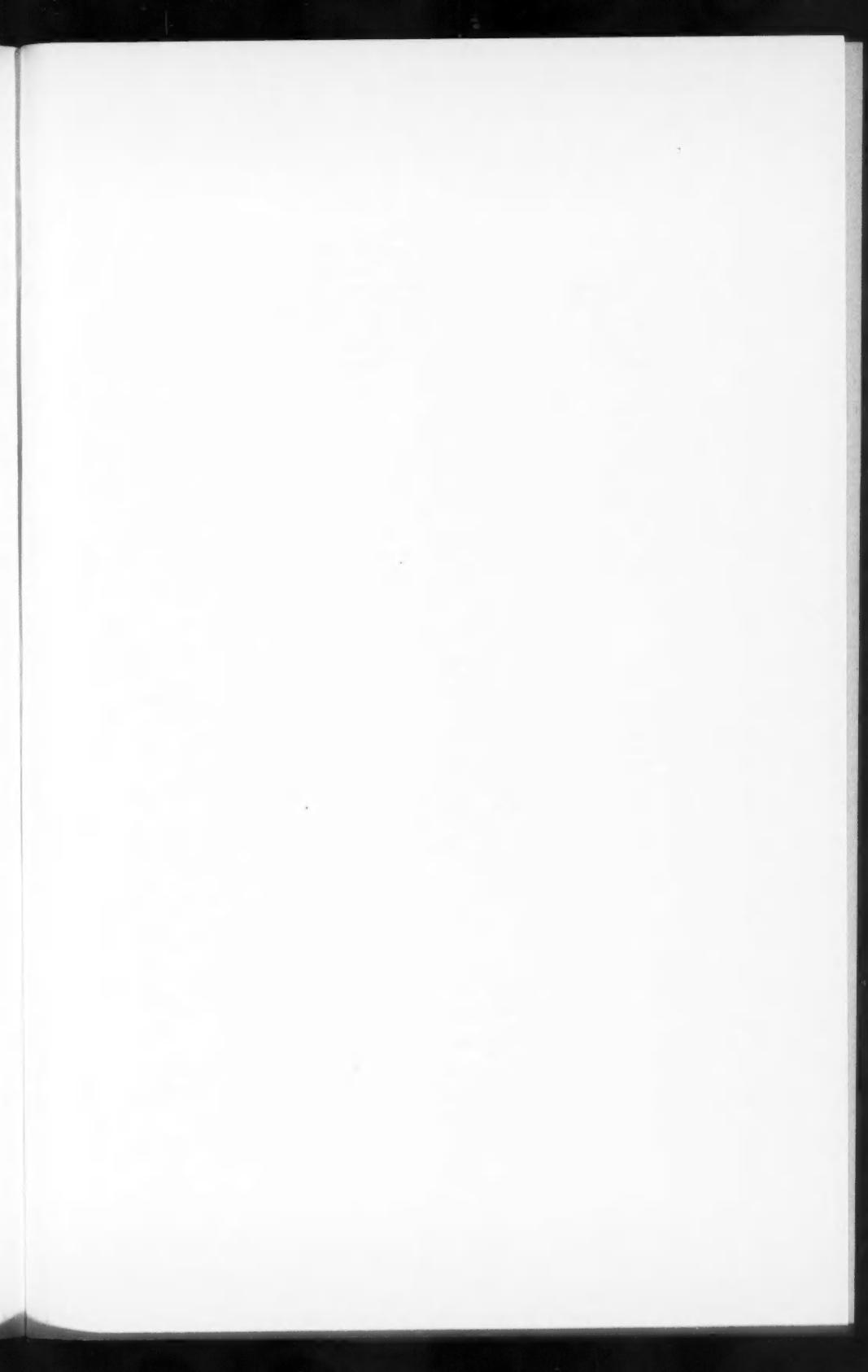
MILLMAN, S., RABI, I. I., and ZACHARIAS, J. R. 1938. *Phys. Rev.* **53**, 384.  
 VARMA, J. and MANDEVILLE, C. E. 1955. *Phys. Rev.* **97**, 977.  
 ZACHARIAS, J. R. 1942. *Phys. Rev.* **61**, 270.

RECEIVED NOVEMBER 17, 1960.  
 DEPARTMENT OF PHYSICS,  
 MCMASTER UNIVERSITY,  
 HAMILTON, ONTARIO.

H. J. KING<sup>2</sup>  
 J. A. CAMERON  
 H. K. EASTWOOD  
 R. G. SUMMERS-GILL

<sup>2</sup>Present address: Goodrich High Voltage Astronautics, Burlington, Mass., U.S.A.







## NOTES TO CONTRIBUTORS

### *Canadian Journal of Physics*

#### MANUSCRIPTS

**General.**—Manuscripts, in English or French, should be typewritten, double spaced, on paper  $8\frac{1}{2} \times 11$  in. The original and one copy are to be submitted. Tables and captions for the figures should be placed at the end of the manuscript. Every sheet of the manuscript should be numbered. Style, arrangement, spelling, and abbreviations should conform to the usage of recent numbers of this journal. Greek letters or unusual signs should be written plainly or explained by marginal notes. Characters to be set in boldface type should be indicated by a wavy line below each character. Superscripts and subscripts must be legible and carefully placed. Manuscripts and illustrations should be carefully checked before they are submitted. Authors will be charged for unnecessary deviations from the usual format and for changes made in the proof that are considered excessive or unnecessary.

**Abstract.**—An abstract of not more than about 200 words, indicating the scope of the work and the principal findings, is required, except in Notes.

**References.**—References should be listed alphabetically by authors' names, unnumbered, and typed after the text. The form of the citations should be that used in current issues of this journal; in references to papers in periodicals, titles should not be given and only initial page numbers are required. The names of periodicals should be abbreviated in the form given in the most recent *List of Periodicals Abstracted by Chemical Abstracts*. All citations should be checked with the original articles and each one referred to in the text by the authors' names and the year.

**Tables.**—Tables should be numbered in roman numerals and each table referred to in the text. Titles should always be given but should be brief; column headings should be brief and descriptive matter in the tables confined to a minimum. Vertical rules should not be used. Numerous small tables should be avoided.

#### ILLUSTRATIONS

**General.**—All figures (including each figure of the plates) should be numbered consecutively from 1 up, in arabic numerals, and each figure referred to in the text. The author's name, title of the paper, and figure number should be written in the lower left corner of the sheets on which the illustrations appear. Captions should not be written on the illustrations.

**Line drawings.**—Drawings should be carefully made with India ink on white drawing paper, blue tracing linen, or co-ordinate paper ruled in blue only; any co-ordinate lines that are to appear in the reproduction should be ruled in black ink. Paper ruled in green, yellow, or red should not be used. All lines must be of sufficient thickness to reproduce well. Decimal points, periods, and stippled dots must be solid black circles large enough to be reduced if necessary. Letters and numerals should be neatly made, preferably with a stencil (do NOT use typewriting) and be of such size that the smallest lettering will be not less than 1 mm high when the figure is reduced to a suitable size. Many drawings are made too large; originals should not be more than 2 or 3 times the size of the desired reproduction. Whenever possible two or more drawings should be grouped to reduce the number of cuts required. In such groups of drawings, or in large drawings, full use of the space available should be made; the ratio of height to width should conform to that of a journal page ( $4\frac{1}{2} \times 7\frac{1}{2}$  in.), but allowance must be made for the captions. The original drawings and one set of clear copies (e.g. small photographs) are to be submitted.

**Photographs.**—Prints should be made on glossy paper, with strong contrasts. They should be trimmed so that essential features only are shown and mounted carefully, with rubber cement, on white cardboard, with no space between those arranged in groups. In mounting, full use of the space available should be made. Photographs are to be submitted in duplicate; if they are to be reproduced in groups one set should be mounted, the duplicate set unmounted.

#### REPRINTS

A total of 100 reprints of each paper, without covers, are supplied free. Additional reprints, with or without covers, may be purchased at the time of publication.

Charges for reprints are based on the number of printed pages, which may be calculated approximately by multiplying by 0.6 the number of manuscript pages (double-spaced typewritten sheets,  $8\frac{1}{2} \times 11$  in.) and including the space occupied by illustrations. Prices and instructions for ordering reprints are sent out with the galley proof.

## Contents

<i>M. A. Clark and J. M. Robson</i> —A polarized neutron beam produced by Bragg reflection from Co-Fe alloy	1
<i>M. A. Clark and J. M. Robson</i> —The electron asymmetry in the beta decay of polarized neutrons	13
<i>F. A. Kaempffer</i> —Representation of space inversion, time reversal, and particle conjugation in quantum field theory	22
<i>C. Collins, D. H. Jelly, and A. G. Matthews</i> —High-frequency radio-wave blackouts at medium and high latitudes during a solar cycle	35
<i>J. M. Daniels, J. C. Giles, and M. A. R. LeBlanc</i> —Nuclear orientation in antiferromagnetic single crystals	53
<i>G. W. Farnell</i> —Elastic waves in trigonal crystals	65
<i>R. F. Millar</i> —Scattering by a grating. I	81
<i>R. F. Millar</i> —Scattering by a grating. II	104
<i>W. G. Nilson</i> —The effect of quenching and neutron irradiation on internal friction of aluminum-5% magnesium alloy	119
<i>R. C. Bourret</i> —Turbulent diffusion in two and three dimensions by the random-walk model with memory	133
<i>H. A. Venables</i> —Maxwell's equations and matrix elements in quantum electrodynamics	141
<i>W. A. Pieczenka, H. E. Petch, and A. B. McLay</i> —An electron spin resonance study of manganese impurity in brucite	145
<i>Guy Lansraus and Germain Boivin</i> —Maximum of the factor of encircled energy	158
<i>J. D. Poll and J. Van Kranendonk</i> —Theory of translational absorption in gases	189
<i>M. J. Y. Clement and D. A. Ramsay</i> —Predissociation in the HNO molecule	205
<i>J. W. C. Johns and D. A. Ramsay</i> —The absorption spectrum and dissociation energy of SH	210
 Notes:	
<i>Peter Rastall</i> —Mach's principle and scalar theories of gravitation	218
<i>A. W. Adey and W. J. Heikkila</i> —Rocket electron density measurements at Fort Churchill, Canada	219
<i>J. R. Prescott and A. S. Rupaal</i> —The specific fluorescence of plastic scintillator NE102	221
 Letters to the Editor:	
<i>S. N. Kalra</i> —Frequency measurement of standard frequency transmissions	228
<i>S. Standil, R. P. Bukata, and F. K. Chin</i> —The sudden increase in the photon component of cosmic radiation on May 4, 1960	229
<i>H. J. King, J. A. Cameron, H. K. Eastwood, and R. G. Summers-Gill</i> —The nuclear spin of indium-115m	230

